### **PCT**

## WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



(51) International Patent Classification 5:		(11) International Publication Number: WO 92/06
C12N 15/54, 9/12, 1/21	A1	(43) International Publication Date: 16 April 1992 (16.0
(22) International Filing Date: 30 September 199	JS91/070	(75) Inventors/Applicants (for US only): GELFAND, David
(30) Priority data:       590,213       28 September 1990 (28.         590,466       28 September 1990 (28.         590,490       28 September 1990 (28.	.09.90) 1	S Third Street, Emeryville, CA 94608 (US).
(60) Parent Application or Grant (63) Related by Continuation US Filed on 28 September 199 (71) Applicant (for all designated States except US CORPORATION [US/US]; 1400 Fifty-Th Emeryville, CA 94608 (US).	S): CETU	(European patent), IT (European patent), JP, LU (European patent), SE (European tent), US.
54) Title: 5' TO 3' EXONUCLEASE MUTATION	is of T	ERMOSTABLE DNA POLYMERASES
57) Abstract		

The present invention relates to thermostable DNA polymerases which exhibit a different level of 5' to 3' exonuclease activity than their respective native polymerases. Particular conserved amino acid domains in thermostable DNA polymerases are mutated or deleted to alter the 5' to 3' exonuclease activity of the polymerases. The present invention also relates to means for isolating and producing such altered polymerases.

#### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	Fi	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greeze	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	JΤ	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CC	Congo	KP	Democratic People's Republic	SE	Sweden
CH	Switzerland		of Korea	SN	Senegal
Ci	Côte d'Ivoire	KR	Republic of Korea	su+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
cs	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE*	Germany	LU	Luxembourg	US	United States of America
DK _	Denmark	MC	Monaco		

<sup>+</sup> Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

WO 92/06200 PCT/US91/07035

5

# 5' TO 3' EXONUCLEASE MUTATIONS OF THERMOSTABLE DNA POLYMERASES

10

#### Cross-Reference to Related Applications

This is a continuation-in-part (CIP) of copending Serial Nos. 590,213, 590,466 and 590,490 all of which 15 were filed on September 28, 1990, and all of which are CIPs of Serial No. 523,394, filed May 15, 1990, which is a CIP of abandoned Serial No. 143,441, filed January 12, 1988, which is a CIP of Serial No. 063,509, filed June 17, 1987, which issued as United States Patent No. 20 4,889,818 and which is a CIP of abandoned Serial No. 899,241, filed August 22, 1986.

This is a also a CIP of Serial No. 746,121 filed August 15, 1991 which is a CIP of: 1) PCT/US90/07641, filed December 21, 1990, which is a CIP of Serial No.

25 585,471, filed September 20, 1990, which is a CIP of Serial No. 455,611, filed December 22, 1989, which is a CIP of Serial No. 143,441, filed January 12, 1988 and its ancestors as described above; and 2) Serial No. 609,157, filed November 2, 1990, which is a CIP of 30 Serial No. 557,517, filed July 24, 1990.

This CIP is also related to the following patent applications:

U.S. Serial No. 523,394, filed May 15, 1990;
U.S. Serial No. 455,967, filed December 22, 1989;
PCT Application No. 91/05571, filed August 6, 1991;
PCT Application No. 91/05753, filed August 13, 1991.

All of the patent applications referenced in this 40 section are incorporated herein by reference.

#### Background of the Invention

#### Field of the Invention

5 The present invention relates to thermostable DNA polymerases which have been altered or mutated such that a different level of 5' to 3' exonuclease activity is exhibited from that which is exhibited by the native enzyme. The present invention also relates to means 10 for isolating and producing such altered polymerases. Thermostable DNA polymerases are useful in many recombinant DNA techniques, especially nucleic acid amplification by the polymerase chain reaction (PCR) self-sustained sequence replication (3SR), and high 15 temperature DNA sequencing.

#### Background Art

Extensive research has been conducted on the 20 isolation of DNA polymerases from mesophilic microorganisms such as E. coli. See, for example, Bessman et al., 1957, J. Biol. Chem. 223:171-177 and Buttin and Kornberg, 1966, <u>J. Biol. Chem.</u> 241:5419-5427. Somewhat less investigation has been made on the 25 isolation and purification of DNA polymerases from thermophiles such Thermus aquaticus, as Thermus thermophilus, Thermotoga maritima, Thermus species sps 17, Thermus species Z05 and Thermosipho africanus. The use of thermostable enzymes to amplify existing 30 nucleic acid sequences in amounts that are large compared to the amount initially present was described in United States Patent Nos. 4,683,195 and 4,683,202, which describe the PCR process, both disclosures of which are incorporated herein by reference. 35 template, nucleoside triphosphates, the appropriate

buffer and reaction conditions, and polymerase are used

in the PCR process, which involves denaturation of target DNA, hybridization of primers, and synthesis of The extension product of each complementary strands. primer becomes a template for the production of the acid sequence. The two patents 5 desired nucleic the polymerase employed disclose that, if thermostable enzyme, then polymerase need not be added after every denaturation step, because heat will not destroy the polymerase activity.

United States Patent No. 4,889,818, European Patent 10 258,017 and PCT Publication No. Publication 89/06691, the disclosures of which are incorporated herein by reference, all describe the isolation and recombinant expression of an ~94 kDa thermostable DNA 15 polymerase from Thermus aquaticus and the use of that T. aquaticus DNA PCR. Although polymerase in polymerase is especially preferred for use in PCR and other recombinant DNA techniques, there remains a need for other thermostable polymerases.

20

#### Summary of the Invention

In addressing the need for other thermostable polymerases, the present inventors found that some thermostable DNA polymerases such as that isolated from thermus aquaticus (Taq) display a 5' to 3' exonuclease or structure-dependent single-stranded endonuclease (SDSSE) activity. As is explained in greater detail below, such 5' to 3' exonuclease activity is undesirable in an enzyme to be used in PCR, because it may limit the amount of product produced and contribute to the plateau phenomenon in the normally exponential accumulation of product. Furthermore, the presence of 5' to 3' nuclease activity in a thermostable DNA polymasseries may contribute to an impaired ability to efficiently generate long PCR products greater than or

equal to 10 kb particularly for G+C-rich targets. In DNA sequencing applications and cycle sequencing applitions, the presence of 5' to 3' nuclease activity may contribute to reduction in desired band intensities and/or generation of spurious or background bands. Finally, the absence of 5' to 3' nuclease activity may facilitate higher sensitivity allelic discrimination in a combined polymerase ligase chain reaction (PLCR) assay.

However, an enhanced or greater amount of 5' to 3' exonuclease activity in a thermostable DNA polymerase may be desirable in such an enzyme which is used in a homogeneous assay system for the concurrent amplification and detection of a target nucleic acid sequence.

15 Generally, an enhanced 5' to 3' exonuclease activity is defined an enhanced rate of exonuclease cleavage or an enhanced rate of nick-translation synthesis or by the displacement of a larger nucleotide fragment before cleavage of the fragment.

20 Accordingly, the present invention was developed to meet the needs of the prior art by providing thermostable DNA polymerases which exhibit altered 5' to 3' exonuclease activity. Depending on the purpose for which the thermostable DNA polymerase will be used, the 25 5' to 3' exonuclease activity of the polymerase may be altered such that a range of 5' to 3' exonuclease activity may be expressed. This range of 5' to 3' exonuclease activity extends from an enhanced activity to a complete lack of activity. Although enhanced 30 activity is useful in certain PCR applications, e. g. a homogeneous assay, as little 5' to 3' exonuclease activity as possible is desired in thermostable DNA polymerases utilized in most other PCR applications.

It was also found that both site directed 35 mutagenesis as well as deletion mutagenesis may result in the desired altered 5' to 3' exonuclease activity in

polymerases of the thermostable DNA the Some mutations which alter the exonuclease invention. activity have been shown to alter the processivity of polymerase. In many applications 5 amplification of moderate sized targets in the presence of a large amount of high complexity genomic DNA) reduced processivity may simplify the optimization of PCRs and contribute to enhanced specificity at high enzyme concentration. Some mutations which eliminate 10 5' to 3' exonuclease activity do not reduce and may enhance the processivity of the thermostable polymerase and accordingly, these mutant enzymes may be preferred in other applications (e.g. generation of long PCR products). Some mutations which eliminate the 15 5' to 3' exonuclease activity simultaneously enhance, relative to the wild type, the thermoresistance of the mutant thermostable polymerase, and thus, these mutant enzymes find additional utility in the amplification of G+C-rich or otherwise difficult to denature targets.

Particular common regions or domains of thermostable DNA polymerase genomes have been identified as preferred sites for mutagenesis to affect the enzyme's 5' to 3' exonuclease. These domains can be isolated and inserted into a thermostable DNA polymerase having 25 none or little natural 5' to 3' exonuclease activity to enhance its activity. Thus, methods of preparing chimeric thermostable DNA polymerases with altered 5' to 3' exonuclease are also encompassed by the present invention.

30

#### Detailed Description of the Invention

The present invention provides DNA sequences and expression vectors that encode thermostable DNA 35 polymerases which have been mutated to alter the

expression of 5' to 3' exonuclease. To facilitate understanding of the invention, a number of terms are defined below.

The terms "cell", "cell line", and "cell culture" can be used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed cells" include the primary transformed cell and cultures derived from that cell without regard to the number of transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

"control The term sequences" refers DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable 20 procaryotes, for example, include promoter, optionally an operator sequence, a ribosome binding site, and possibly other sequences. Eucaryotic cells utilize promoters, are known to polyadenylation signals, and enhancers.

25 term "expression system" The refers DNA sequences containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed with these sequences are capable producing the encoded proteins. To 30 transformation, the expression system may be included on a vector; however, the relevant DNA may also be integrated into the host chromosome.

The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for 35 the production of a recoverable bioactive polypeptide or precursor. The polypeptide can be encoded by a full

length coding sequence or by any portion of the coding sequence so long as the enzymatic activity is retained.

The term "operably linked" refers to the positioning of the coding sequence such that control sequences will function to drive expression of the protein encoded by the coding sequence. Thus, a coding sequence "operably linked" to control sequences refers to a configuration wherein the coding sequences can be expressed under the direction of a control sequence.

The term "mixture" as it relates to mixtures containing thermostable polymerases refers to a collection of materials which includes a desired thermostable polymerase but which can also include other proteins. If the desired thermostable polymerase is derived from recombinant host cells, the other proteins will ordinarily be those associated with the host. Where the host is bacterial, the contaminating proteins will, of course, be bacterial proteins.

The term "non-ionic polymeric detergents" refers to 20 surface-active agents that have no ionic charge and that are characterized for purposes of this invention, by an ability to stabilize thermostable polymerase enzymes at a pH range of from about 3.5 to about 9.5, preferably from 4 to 8.5.

"oligonucleotide" as used herein The term 25 a molecule comprised of two or as ribonucleotides, deoxyribonucleotides or more than three, and usually more than ten. The exact size will depend on many factors, which in turn depends ultimate function oruse 30 on the The oligonucleotide may be derived oligonucleotide. synthetically or by cloning.

The term "primer" as used herein refers to an oligonucleotide which is capable of acting as a point 35 of initiation of synthesis when placed under conditions in which primer extension is initiated. An

oligonucleotide "primer" may occur naturally, as in a purified restriction digest or be produced synthetically. Synthesis of a primer extension product which is complementary to a nucleic acid strand is 5 initiated in the presence of four different nucleoside triphosphates and a thermostable polymerase enzyme in an appropriate buffer at a suitable temperature. "buffer" includes cofactors (such as divalent metal and salt (to provide the appropriate ionic 10 strength), adjusted to the desired pH.

A primer is single-stranded for maximum efficiency in amplification, but may alternatively double-stranded. If double-stranded, the primer first treated to separate its strands before being used 15 to prepare extension products. The primer is usually oligodeoxyribonucleotide. The primer must sufficiently long to prime the synthesis of extension products in the presence of the polymerase enzyme. exact length of a primer will depend on many factors, 20 such as source of primer and result desired, and the reaction temperature must be adjusted depending on primer length and nucleotide sequence to ensure proper annealing of primer to template. Depending on the complexity of the target sequence, an oligonucleotide 25 primer typically contains 15 to 35 nucleotides. primer molecules generally require lower temperatures to form sufficiently stable complexes with template.

primer is selected to be "substantially" complementary to a strand of specific sequence of the 30 template. A primer must be sufficiently complementary hybridize with a template strand for primer elongation to occur. A primer sequence need not reflect the exact sequence of the template. example, a non-complementary nucleotide fragment may be 35 attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridize and thereby form a template primer complex for synthesis of the extension product of the primer.

The terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes which 10 cut double-stranded DNA at or near a specific nucleotide sequence.

The term "thermostable polymerase enzyme" refers to an enzyme which is stable to heat and is heat resistant and catalyzes (facilitates) combination of the nucleotides in the proper manner to form primer extension products that are complementary to a template nucleic acid strand. Generally, synthesis of a primer extension product begins at the 3' end of the primer and proceeds in the 5' direction along the template strand, until synthesis terminates.

In order to further facilitate understanding of the invention, specific thermostable DNA polymerase enzymes are referred to throughout the specification to exemplify the broad concepts of the invention, and 25 these references are not intended to limit the scope of the invention. The specific enzymes which are frequently referenced are set forth below with a common abbreviation which will be used in the specification and their respective nucleotide and amino acid Sequence 30 ID numbers.

Thermostable DNA Polymerase	Common <u>Abbr.</u>	SEQ. ID NO:			
35 <u>Thermus</u> <u>aquaticus</u>	Tag	SEQ ID NO:1	(nuc)		
·		SEQ ID NO:2	(a.a.)		

WO 92/06200 PCT/US91/07035

-10-

			NO:3	(nuc)
	SEQ	ID	NO:4	(a.a.)
sps17 <u>Tsps17</u>	SEQ	ID	NO:5	(nuc)
	SEQ	ID	NO:6	(a.a.)
Z05 <u>TZ05</u>	SEQ	ID	NO:7	(nuc)
	SEQ	ID	NO:8	(a.a.)
ilus Tth	SEQ	ID	NO:9	(nuc)
	SEQ	ID	NO:10	(a.a.)
canus <u>Taf</u>	SEQ	ID	NO:11	(nuc)
	SEQ	ID	NO:12	(a.a.)
	Z05 <u>TZ05</u> ilus <u>Tth</u>	sps17 Tsps17 SEQ  SEQ  ZO5 TZO5 SEQ  SEQ  ilus Tth SEQ  SEQ  SEQ  SEQ	sps17 Tsps17 SEQ ID SEQ ID Z05 TZ05 SEQ ID SEQ ID ilus Tth SEQ ID SEQ ID SEQ ID	SEQ ID NO:6 Z05 TZ05 SEQ ID NO:7 SEQ ID NO:8 ilus Tth SEQ ID NO:9 SEQ ID NO:10

As summarized above, the present invention relates to thermostable DNA polymerases which exhibit altered 5' to 3' exonuclease activity from that of the native polymerase. Thus, the polymerases of the invention 25 exhibit either an enhanced 5' to 3' exonuclease activity or an attenuated 5' to 3' exonuclease activity from that of the native polymerase.

# Thermostable DNA Polymerases With Attenuated 30 5' to 3' Exonuclease Activity

DNA polymerases often possess multiple functions. In addition to the polymerization of nucleotides <u>E</u>. <u>coli</u> DNA polymerase I (pol I), for example, catalyzes 35 the pyrophosphorolysis of DNA as well as the hydrolysis of phosphodiester bonds. Two such hydrolytic activities have been characterized for pol I; one is a 3' to 5' exonuclease activity and the other a 5' to 3' exonuclease activity. The two exonuclease activities 40 are associated with two different domains of the pol I molecule. However, the 5' to 3' exonuclease activity of pol I differs from that of thermostable DNA

PCT/US91/07035

polymerases in that the 5' to 3' exonuclease activity of thermostable DNA polymerases has stricter structural requirements for the substrate on which it acts.

An appropriate and sensitive assay for the 5' to 3' 5 exonuclease activity of thermostable DNA polymerases takes advantage of the discovery of the structural requirement of the activity. An important feature of the design of the assay is an upstream oligonucleoside primer which positions the polymerase appropriately for of a labeled downstream cleavage oligonucleotide probe. For an assay of polymerizationexonuclease activity (i.e., an independent absence of deoxynucleoside performed in the triphosphates) the probe must be positioned such that 15 the region of probe complementary to the template is immediately adjacent to the 3'-end of the primer. Additionally, the probe should contain at least one, but preferably 2-10, or most preferably 3-5 nucleotides at the 5'-end of the probe which are not complementary 20 to the template. The combination of the primer and probe when annealed to the template creates a double stranded structure containing a nick with a 3'-hydroxyl 5' of the nick, and a displaced single strand 3' of the Alternatively, the assay can be performed as a 25 polymerization-dependent reaction, in which case each deoxynucleoside triphosphate should be included at a concentration of between 1 µM and 2 mM, preferably between 10 µM and 200 µM, although limited dNTP addition (and thus limited dNTP inclusion) may be 30 involved as dictated by the template sequence. the assay is performed in the presence of dNTPs, the necessary structural requirements are an upstream oligonucleotide primer to direct the synthesis of the complementary strand of the template by the polymerase, 35 and a labeled downstream oligonucleotide probe which will be contacted by the polymerase in the process of

WO 92/06200

extending the upstream primer. An example of a polymerization-independent thermostable DNA polymerase 5' to 3' exonuclease assay follows.

synthetic 3' phosphorylated oligonucleotide 5 probe (phosphorylated to preclude polymerase extension) (GATCGCTGCGCGTAACCACCACACCCGCCGCGCp) NO:13) (100 pmol) was <sup>32</sup>P-labeled at the 5' end with gamma-[32P] ATP (3000 Ci/mmol) and T4 polynucleotide The reaction mixture was extracted with 10 phenol:chloroform:isoamyl alcohol, followed by ethanol The <sup>32</sup>P-labeled oligonucleotide probe precipitation. of redissolved in 100 μl TE buffer, unincorporated ATP was removed by gel filtration chromatography on a Sephadex G-50 spin column. 15 pmol of <sup>32</sup>P-labeled BW33 probe, was annealed to 5 pmol of single-strand M13mp10w DNA, in the presence of 5 pmol of the synthetic oligonucleotide primer BW37 (GCGCTAGGGCGCTGGCAAGTGTAGCGGTCA) (SEQ ID NO:14) in a 100 µl reaction containing 10 mM Tris-HCl (pH 8.3), 20 50 mM KCl, and 3 mM MgCl2. The annealing mixture was heated to 95°C for 5 minutes, cooled to 70°C over 10 minutes, incubated at 70°C for an additional minutes, and then cooled to 25°C over a 30 minute period in a Perkin-Elmer Cetus DNA Thermal Cycler. 25 Exonuclease reactions containing 10 µl of the annealing mixture were pre-incubated at 70°C for 1 minute. Thermostable DNA polymerase enzyme (approximately 0.01 to 1 unit of DNA polymerase activity, or 0.0005 to 0.05 pmol of enzyme) was added in a 2.5 µl volume to the 30 pre-incubation reaction, and the reaction mixture was incubated at 70°C. Aliquots (5 µl) were removed after 1 minute and 5 minutes, and stopped by the addition of The reaction products were 1 ul of 60 mM EDTA. analyzed by homochromatography and exonuclease activity following 35 was quantified autoradiography. Chromatography was carried out in a homochromatography

mix containing 2% partially hydrolyzed yeast RNA in 7M urea on Polygram CEL 300 DEAE cellulose thin layer chromatography plates. The presence of 5' to 3' exonuclease activity results in the generation of small <sup>32</sup>p-labeled oligomers, which migrate up the TLC plate, and are easily differentiated on the autoradiogram from undegraded probe, which remains at the origin.

exonuclease 3' activity the 5′ to The terminal DNA polymerases excises thermostable 10 regions of double-stranded DNA releasing 5'-mono- and oligonucleotides in a sequential manner. The preferred substrate for the exonuclease is displaced singlestranded DNA, with hydrolysis of the phosphodiester bond occurring between the displaced single-stranded double-helical DNA. The 15 DNA and the exonuclease cleavage site is a phosphodiester bond in Thus, the exonuclease the double helical region. described as better can be activity single-stranded endonuclease structure-dependent 20 (SDSSE).

Many thermostable polymerases exhibit this 5' to 3' exonuclease activity, including the DNA polymerases of Tag, Tma, Tsps17, TZ05, Tth and Taf. When thermostable polymerases which have 5' to 3' exonuclease activity 25 are utilized in the PCR process, a variety undesirable results have been observed including a limitation of the amount of product produced, impaired ability to generate long PCR products or regions containing significant amplify 30 structure, the production of shadow bands or the attenuation in signal strength of desired termination bands during DNA sequencing, the degradation of the 5'-end of oligonucleotide primers in the context of complex, nickprimer-template double-stranded

translation synthesis during oligonucleotide-directed mutagenesis and the degradation of the RNA component of RNA: DNA hybrids.

The limitation of the amount of PCR product 5 produced is attributable to a plateau phenomenon in the otherwise exponential accumulation of product. Such a plateau phenomenon occurs in part because 5' to 3' exonuclease activity causes the hydrolysis or cleavage of phosphodiester bonds when a polymerase with 5' to 3' 10 exonuclease activity encounters a forked structure on a PCR substrate.

Such forked structures commonly exist in certain Gand C-rich DNA templates. The cleavage of these
phosphodiester bonds under these circumstances is
15 undesirable as it precludes the amplification of
certain G- and C-rich targets by the PCR process.
Furthermore, the phosphodiester bond cleavage also
contributes to the plateau phenomenon in the generation
of the later cycles of PCR when product strand
20 concentration and renaturation kinetics result in
forked structure substrates.

In the context of DNA sequencing, the 5' to 3' exonuclease activity of DNA polymerases is again a hinderance with forked structure templates because the 25 phosphodiester bond cleavage during the DNA extension reactions results in "false stops". These "false stops" in turn contribute to snadow bands, and in extreme circumstances may result in the absence of accurate and interpretable sequence data.

When utilized in a PCR process with double-stranded primer-template complex, the 5' to 3' exonuclease activity of a DNA polymerase may result in the degradation of the 5'-end of the oligonucleotide primers. This activity is not only undesirable in PCR, but also in second-strand cDNA synthesis and sequencing processes.

During optimally efficient oligonucleotide-directed mutagenesis processes, the DNA polymerase which is utilized must not have strand-displacement synthesis and/or nick-translation capability. Thus, the presence of 5' to 3' exonuclease activity in a polymerase used for oligonucleotide-directed mutagenesis is also undesirable.

Finally, the 5' to 3' exonuclease activity of polymerases generally also contains an inherent RNase H 10 activity. However, when the polymerase is also to be used as a reverse transcriptase, as in a PCR process including an RNA: DNA hybrid, such an inherent RNase H activity may be disadvantageous.

Thus, one aspect of this invention involves the 15 generation of thermostable DNA polymerase mutants displaying greatly reduced, attenuated or completely eliminated 5' to 3' exonuclease activity. Such mutant thermostable DNA polymerases will be more suitable and desirable for use in processes such as PCR, second-20 strand cDNA synthesis, sequencing and oligonucleotide-directed mutagenesis.

The production of thermostable DNA polymerase mutants with attenuated or eliminated 5' to 3' exonuclease activity may be accomplished by processes such as site-directed mutagenesis and deletion mutagenesis.

For example, a site-directed mutation of G to A in the second position of the codon for Gly at residue 46 in the <u>Taq</u> DNA polymerase amino acid sequence (i.e. 30 mutation of G(137) to (A) in the DNA sequence has been found to result in an approximately 1000-fold reduction of 5' to 3' exonuclease activity with no apparent change in polymerase activity, processivity or extension rate. This site-directed mutation of the <u>Tag</u> 35 DNA polymerase nucleotide sequence results in an amino acid change of Gly (46) to Asp.

Glycine 46 of <u>Taq</u> DNA polymerase is conserved in <u>Thermus</u> species sps17 DNA polymerase, but is located at residue 43, and the same Gly to Asp mutation has a similar effect on the 5' to 3' exonuclease activity of <u>Tsps17</u> DNA polymerase. Such a mutation of the conserved Gly of <u>Tth</u> (Gly 46), <u>Tz05</u> (Gly 46), <u>Tma</u> (Gly 37) and <u>Taf</u> (Gly 37) DNA polymerases to Asp also has a similar attenuating effect on the 5' to 3' exonuclease activities of those polymerases.

- Tsps17 Gly 43, Tth Gly 46, TZ05 Gly 46, Tma Gly 37 10 and  $\underline{\text{Taf}}$  Gly 37 are also found in a conserved A(V/T)YG (SEQ ID NO:15) sequence domain, and changing the glycine to aspartic acid within this conserved sequence domain of any polymerase is also expected to attenuate 15 5' to 3' exonuclease activity. Specifically, Tsps17 Gly 43, Tth Gly 46, TZ05 Gly 46, and Taf Gly 37 share the AVYG sequence domain, and Tma Gly 37 is found in the ATYG domain. Mutations of glycine to aspartic acid in other thermostable DNA polymerases containing the 20 conserved A(V/T)YG (SEQ ID NO:15) domain can be accomplished utilizing the same principles techniques used for the site-directed mutagenesis of Tag polymerase. Exemplary of such site-directed mutagenesis techniques are Example 5 of U.S. Serial 25 No. 523,394, filed May 15, 1990, Example 4 of Attorney Docket No. 2583.1 filed September 27, 1991, Examples 4 and 5 of U.S. Serial No. 455,967, filed December 22, 1989 and Examples 5 and 8 of PCT Application No. 91/05753, filed August 13, 1991.
- 30 Such site-directed mutagenesis is generally accomplished by site-specific primer-directed mutagenesis. This technique is now standard in the art, and is conducted using a synthetic oligonucleotide primer complementary to a single-stranded phage DNA to 35 be mutagenized except for limited mismatching, representing the desired mutation. Briefly,

synthetic oligonucleotide is used as a primer to direct synthesis of a strand complementary to the phasmid or phage, and the resulting double-stranded DNA is transformed into a phage-supporting host bacterium. Cultures of the transformed bacteria are plated in top

5 Cultures of the transformed bacteria are plated in top agar, permitting plaque formation from single cells that harbor the phage or plated on drug selective media for phasmid vectors.

Theoretically, 50% of the new plaques will contain 10 the phage having, as a single strand, the mutated form; 50% will have the original sequence. The plaques are tranferred to nitrocellulose filters and the "lifts" hybridized with kinased synthetic primer temperature that permits hybridization of an exact 15 match, but at which the mismatches with the original sufficient strand to prevent hybridization. Plaques that hybridize with the probe are then picked and cultured, and the DNA is recovered.

In the constructions set forth below, 20 ligations for plasmid construction are confirmed by first transforming E. coli strains DG98, DG101, DG116, or other suitable hosts, with the ligation mixture. Successful transformants are selected by ampicillin, tetracycline or other antibiotic resistance or using 25 other markers, depending on the mode of plasmid construction, as is understood in the art. from the transformants are then prepared according to the method of Clewell, D.B., et al., Proc. Natl. Acad. Sci. (USA) (1969) <u>62</u>:1159, optionally following 30 chloramphenicol amplification (Clewell, D.B., Bacteriol. (1972) 110: 667). The isolated DNA analyzed by restriction and/or sequenced by the dideoxy method of Sanger, F., et al., Proc. Natl. Acad. Sci. (USA) (1977) 74:5463 as further described by Messing,

et al., <u>Nucleic Acids Res.</u> (1981) <u>9</u>:309, or by the method of Maxam, et al., <u>Methods in Enzymology</u> (1980) <u>65</u>:499.

For cloning and sequencing, and for expression of 5 constructions under control of most <u>lac</u> or P<sub>L</sub> promoters, <u>E. coli</u> strains DG98, DG101, DG116 were used as the host. For expression under control of the P<sub>L</sub>N<sub>RBS</sub> promoter, <u>E. coli</u> strain K12 MC1000 lambda lysogen, N<sub>7</sub>N<sub>53</sub>cI857 SusP<sub>80</sub>, ATCC 39531 may be used.

10 Exemplary hosts used herein for expression of the thermostable DNA polymerases with altered 5' to 3' exonuclease activity are <u>E. coli</u> DG116, which was deposited with ATCC (ATCC 53606) on April 7, 1987 and <u>E. coli</u> KB2, which was deposited with ATCC (ATCC 53075)

15 on March 29, 1985.

For M13 phage recombinants, <u>E. coli</u> strains susceptible to phage infection, such as <u>E. coli</u> K12 strain DG98, are employed. The DG98 strain has been deposited with ATCC July 13, 1984 and has accession 20 number 39768.

Mammalian expression can be accomplished in COS-7 COS-A2, CV-1, and murine cells, and insect cell-based expression in <u>Spodoptera frugipeida</u>.

The thermostable DNA polymerases of the present invention are generally purified from <u>E</u>. <u>coli</u> strain DG116 containing the features of plasmid pLSG33. The primary features are a temperature regulated promoter (λ P<sub>L</sub>, promoter), a temperature regulated plasmid vector, a positive retro-regulatory element (PRE) (see 30 U.S. 4,666,848, issued May 19, 1987), and a modified form of a thermostable DNA polymerase gene. As described at page 46 of the specification of U.S patent application Serial No. 455,967, pLSG33 was prepared by ligating the <u>Nde</u>I-<u>Bam</u>HI restriction fragment of pLSG24 into expression vector pDG178. The resulting plasmids are ampicillin resistant and capable of expressing 5'

to 3' exonuclease deficient forms of the thermostable DNA polymerases of the present invention. flask for a 10 liter fermentation contains tryptone (20 q/l), yeast extract (10 g/l), NaCl (10 g/l) and 0.005% 5 ampicillin. The seed flask is inoculated from colonies from an agar plate, or a frozen glycerol culture stock can be used. The seed is grown to between 0.5 and 1.0 The volume of seed culture inoculated O.D.  $(A_{680})$ . into the fermentation is calculated such that the final of bacteria will be 10 concentration The 10 liter growth medium contained weight/liter. 25 mM  $\mathrm{KH_2PO_4}$ , 10 mM  $\mathrm{(NH_4)_2}$   $\mathrm{SO_4}$ , 4 mM sodium citrate, 0.4 mM FeCl<sub>2</sub>, 0.04 mM ZnCl<sub>2</sub>, 0.03 mM CoCl<sub>2</sub>, 0.03 mM The following sterile  $CuCl_2$ , and 0.03 mM  $H_3BO_3$ . 4 mM MgSO<sub>4</sub>, 20 g/l glucose, 15 components are added: 20 mg/l thiamine-HCl and 50 mg/l ampicillin. was adjusted to 6.8 with NaOH and controlled during the Glucose is continually fermentation by added NH4OH. added during the fermentation by coupling to NH4OH Foaming is controlled by the addition of 20 addition. polypropylene glycol as necessary, as an anti-foaming Dissolved oxygen concentration is maintained at agent. 40%.

The fermentation is inoculated as described above 25 and the culture is grown at 30°C until an optical density of 21  $(A_{680})$  is reached. The temperature is then raised to 37°C to induce synthesis of the desired Growth continues for eight hours after polymerase. and the cells are then harvested induction, 30 concentration using cross flow filtration followed by The resulting cell paste is frozen at centrifugation. -70°C and yields about 500 grams of cell paste. steps all purification indicated, otherwise conducted at 4°C.

A portion of the frozen (-70°C) <u>E. coli</u> K12 strain DG116 harboring plasmid pLSG33 or other suitable host

as described above is warmed overnight to -20°C. the cell pellet the following reagents are added: 1 volume of 2X TE (100 mM Tris-HCl, pH 7.5, 20 mM EDTA), 1 mg/ml leupeptin and 144 mM PMSF (in dimethyl 5 formamide). The final concentration of leupeptin was 1 µg/ml and for PMSF, 2.4 mM. Preferably, dithiothreitol (DTT) is included in TE to provide a final concentration of 1 mm DTT. The mixture is homogenized at low speed in a blender. All glassware 10 is baked prior to use, and solutions used in the purification are autoclaved, if possible, prior to The cells are lysed by passage twice through a Microfluidizer at 10,000 psi.

The lysate is diluted with 1X TE containing 1 mM 15 DTT to a final volume of 5.5X cell wet weight. Leupeptin is added to 1  $\mu$ g/ml and PMSF is added to 2.4 mM. The final volume (Fraction I) is approximately 1540 ml.

Ammonium sulfate is gradually added to 0.2 M (26.4 20 g/l) and the lysate stirred. Upon addition of ammonium sulfate, a precipitate forms which is removed prior to polyethylenimine (PEI) precipitation described below. The ammonium sulfate precipitate is removed by centrifugation of the suspension at 15,000 -25 20,000 xg in a JA-14 rotor for 20 minutes. supernatant is decanted and retained. sulfate supernatant is then stirred on a heating plate until the supernatant reaches 75°C and then is placed in a 77°C bath and held there for 15 minutes with 30 occasional stirring. The supernatant is then cooled in an ice bath to 20°C and a 10 ml aliquot is removed for PEI titration.

PEI titration and agarose gel electrophoresis are used to determine that 0.3% PEI (commercially available 35 from BDH as PolyminP) precipitates ~90% of the macromolecular DNA and RNA, i.e., no DNA band is

WO 92/06200 PCT/US91/07035

visible on an ethidium bromide stained agarose gel after treatment with PEI. PEI is added slowly with stirring to 0.3% from a 10% stock solution. The PEI treated supernatant is centrifuged at 10,000 RPM 5 (17,000 xg) for 20 minutes in a JA-14 rotor. The supernatant is decanted and retained. The volume (Fraction II) is approximately 1340 ml.

Fraction II is loaded onto a 2.6 x 13.3 cm (71 ml) phenyl sepharose CL-4B (Pharmacia-LKB) column following 10 equilibration with 6 to 10 column volumes of containing 0.2 M ammonium sulfate. Fraction II is then loaded at a linear flow rate of 10 cm/hr. rate is 0.9 ml/min. The column is washed with 3 column volumes of the equilibration buffer and then with 2 15 column volumes of TE to remove contaminating non-DNA polymerase proteins. The recombinant thermostable DNA polymerase is eluted with 4 column volumes of 2.5 M urea in TE containing 20% ethylene glycol. identified polymerase containing fractions are DNA polymerase activity 20 optical absorption (A280), assay and SDS-PAGE according to standard procedures. Peak fractions are pooled and filtered through a 0.2 micron sterile vacuum filtration apparatus. (Fraction III) is approximately 195 ml. The resin is according the 25 equilibrated to and recycled manufacturer's recommendations.

A 2.6 x 1.75 cm (93 ml) heparin sepharose C1-6B column (Pharmacia-LKB) is equilibrated with 6-10 column volumes of 0.05 M KCl, 50 mM Tris-HCl, pH 7.5, 0.1 mM 30 EDTA and 0.2% Tween 20, at 1 column volume/hour. Preferably, the buffer contains 1 mM DTT. The column is washed with 3 column volumes of the equilibration buffer. The desired thermostable DNA polymerase of the invention is eluted with a 10 column volume linear gradient of 50-750 mM KCl gradient in the same buffer. Fractions (one-tenth column volume) are collected in

sterile tubes and the fractions containing the desired thermostable DNA polymerase are pooled (Fraction IV, volume 177 ml).

Fraction IV is concentrated to 10 ml on an Amicon 5 YM30 membrane. For buffer exchange, diafiltration is done 5 times with 2.5X storage buffer (50 mM Tris-HCl, pH 7.5, 250 mM KCl, 0.25 mM EDTA 2.5 mM DTT and 0.5% Tween-20) by filling the concentrator to 20 ml and concentrating the volumes to 10 ml each time. The 10 concentrator is emptied and rinsed with 10 ml 2.5X storage buffer which is combined with the concentrate to provide Fraction V.

Anion exchange chromatography is used to remove The procedure is conducted residual 15 biological safety hood and sterile techniques are A Waters Sep-Pak plus QMA cartridge with a 0.2 micron sterile disposable syringe tip filter unit is equilibrated with 30 ml of 2.5% storage buffer using a syringe at a rate of about 5 drops per second. Using a 20 disposable syringe, Fraction V is passed through the cartridge at about 1 drop/second and collected in a sterile tube. The cartridge is flushed with 5 ml of 2.5 ml storage buffer and pushed dry with air. eluant is diluted 1.5 X with 80% glycerol and stored at The resulting final Fraction IV pool contains active thermostable DNA polymerase with altered 5' to 3' exonuclease activity.

In addition to site-directed mutagenesis of a nucleotide sequence, deletion mutagenesis techniques 30 may also be used to attenuate the 5' to 3' exonuclease activity of a thermostable DNA polymerase. One example of such a deletion mutation is the deletion of all amino terminal amino acids up to and including the glycine in the conserved A(V/T)YG (SEQ ID NO:15) domain 35 of thermostable DNA polymerases.

A second deletion mutation affecting 5' to 3' exonuclease activity is a deletion up to Ala 77 in Tag This amino acid (Ala 77) has been DNA polymerase. identified as the amino terminal amino acid in an 5 approximately 85.5 kDa proteolytic product of Tag DNA product has proteolytic polymerase. This Tag DNA polymerase several native identified in preparations and the protein appears to be stable. Since such a deletion up to Ala 77 includes Gly 46, it 10 will also affect the 5' to 3' exonuclease activity of Tag DNA polymerase.

However, a deletion mutant beginning with Ala 77 added advantage over a deletion has the beginning with phenylalanine 47 in that the proteolytic 15 evidence suggests that the peptide will remain stable. Furthermore, Ala 77 is found within the sequence HEAYG (SEQ ID NO:16) 5 amino acids prior to the sequence YKA in <u>Taq</u> DNA polymerase. A similar sequence motif HEAYE (SEQ ID NO:17) is found in Tth DNA polymerase, TZ05 DNA The alanine is 5 20 polymerase and <u>Tsps17</u> DNA polymerase. amino acids prior to the conserved motif YKA. amino acids in the other exemplary thermostable DNA polymerases which correspond to Tag Ala 77 are Tth Ala 78,  $\underline{\text{TZ05}}$  Ala 78,  $\underline{\text{Tsps17}}$  Ala 74,  $\underline{\text{Tma}}$  Leu 72 and  $\underline{\text{Taf}}$  Ile A deletion up to the alanine or corresponding amino acid in the motif HEAY(G/E) (SEQ ID NO:16 or SEQ <u>Thermus</u> species thermostable NO:17) in a polymerase containing this sequence will attenuate its 3 ′ 5′ exonuclease activity. The 3*'* 30 exonuclease motif YKA is also conserved in Tma DNA polymerase (amino acids 76-78) and Taf DNA polymerase (amino acids 77-79). In this thermostable polymerase family, the conserved motif (L/I)LET (SEQ ID NO:18) Taf DNA polymerase immediately proceeds the YKA motif. 35 Ile 73 is 5 residues prior to this YKA motif while TMA DNA polymerase Leu 72 is 5 residues prior to the YKA

motif. A deletion of the Leu or Ile in the motif (L/I)LETYKA (SEQ ID NO:19) in a thermostable DNA polymerase from the <u>Thermotoga</u> or <u>Thermosipho</u> genus will also attenuate 5' to 3' exonuclease activity.

Thus, a conserved amino acid sequence which defines the 5' to 3' exonuclease activity of DNA polymerases of the <u>Thermus</u> genus as well as those of <u>Thermotoga</u> and <u>Thermosipho</u> has been identified as (I/L/A)X<sub>3</sub>YKA (SEQ ID NO:20), wherein X<sub>3</sub> is any sequence of three amino acids. Therefore, the 5' to 3' exonuclease activity of thermostable DNA polymerases may also be altered by mutating this conserved amino acid domain.

Those of skill in the art recognize that when such a deletion mutant is to be expressed in recombinant 15 host cells, a methionine codon is usually placed at the 5' end of the coding sequence, so that the amino terminal sequence of the deletion mutant protein would be MET-ALA in the Thermus genus examples above.

The preferred techniques for performing deletion 20 mutations involve utilization of known restriction sites on the nucleotide sequence of the thermostable DNA polymerase. Following identification of the particular amino acid or amino acids which are to be deleted, a restriction site is identified which when 25 cleaved will cause the cleavage of the target DNA sequence at a position or slightly 3' distal to the position corresponding to the amino acid or domain to be deleted, but retains domains which code for other properties of the polymerase which are desired.

30 Alternatively, restriction sites on either side (5' or 3') of the sequence coding for the target amino acid or domain may be utilized to cleave the sequence. However, a ligation of the two desired portions of the sequence will then be necessary. This ligation may be 35 performed using techniques which are standard in the art and exemplified in Example 9 of Serial No. 523,394,

filed May 15, 1990, Example 7 of PCT Application No. 91/05753, filed August 13, 1991 and Serial No. 590,490, filed September 28, 1990, all of which are incorporated herein by reference.

Another technique for achieving a deletion mutation of the thermostable DNA polymerase is by utilizing the PCR mutagenesis process. In this process, primers are prepared which incorporate a restriction site domain and optionally a methionine codon if such a codon is 10 not already present. Thus, the product of the PCR with may be digested with an appropriate this primer restriction enzyme to remove the domain which codes for 5' to 3' exonuclease activity of the enzyme. two remaining sections of the product are ligated to 15 form the coding sequence for a thermostable DNA polymerase lacking 5' to 3' exonuclease activity. Such coding sequences can be utilized as expression vectors appropriate host cells to produce the 3' 5′ to lacking polymerase thermostable DNA 20 exonuclease activity.

In addition to the <u>Tag</u> DNA polymerase mutants with reduced 5' to 3' exonuclease activity, it has also been found that a truncated <u>Tma</u> DNA polymerase with reduced 5' to 3' exonuclease activity may be produced by recombinant techniques even when the complete coding sequence of the <u>Tma</u> DNA polymerase gene is present in an expression vector in <u>E. coli</u>. Such a truncated <u>Tma</u> DNA polymerase is formed by translation starting with the methionine codon at position 140. Furthermore, recombinant means may be used to produce a truncated polymerase corresponding to the protein produced by initiating translation at the methionine codon at position 284 of the <u>Tma</u> coding sequence.

The <u>Tma</u> DNA polymerase lacking amino acids 1 though 35 139 (about 86 kDa), and the <u>Tma</u> DNA polymerase lacking amino acids 1 through 283 (about 70 kDa) retain

polymerase activity but have attenuated 5' to 3' exonuclease activity. An additional advantage of the 70 kDa <u>Tma</u> DNA polymerase is that it is significantly more thermostable than native <u>Tma</u> polymerase.

Thus, it has been found that the entire sequence of the intact <a href="Tma">Tma</a> DNA polymerase I enzyme is not required for activity. Portions of the <a href="Tma">Tma</a> DNA polymerase I coding sequence can be used in recombinant DNA techniques to produce a biologically active gene 10 product with DNA polymerase activity.

Furthermore, the availability of DNA encoding the Tma DNA polymerase sequence provides the opportunity to modify the coding sequence so as to generate mutein (mutant protein) forms also having DNA polymerase activity but with attenuated 5' to 3' exonuclease activity. The amino(N)-terminal portion of the Tma DNA polymerase is not necessary for polymerase activity but rather encodes the 5' to 3' exonuclease activity of the protein.

- Thus, using recombinant DNA methodology, one can delete approximately up to one-third of the N-terminal coding sequence of the <u>Tma</u> gene, clone, and express a gene product that is quite active in polymerase assays but, depending on the extent of the deletion, has no 5' to 3' exonuclease activity. Because certain N-terminal shortened forms of the polymerase are active, the gene constructs used for expression of these polymerases can include the corresponding shortened forms of the coding sequence.
- In addition to the N-terminal deletions, individual amino acid residues in the peptide chain of Tma DNA polymerase or other thermostable DNA polymerases may be modified by oxidation, reduction, or other derivation, and the protein may be cleaved to obtain fragments that retain polymerase activity but have attenuated 5' to 3' exonuclease activity. Modifications to the primary

structure of the <u>Tma</u> DNA polymerase coding sequence or the coding sequences of other thermostable DNA polymerases by deletion, addition, or alteration so as to change the amino acids incorporated into the thermostable DNA polymerase during translation of the mRNA produced from that coding sequence can be made without destroying the high temperature DNA polymerase activity of the protein.

Another technique for preparing thermostable DNA 10 polymerases containing novel properties such as reduced or enhanced 5' to 3' exonuclease activity is a "domain of for the construction shuffling" technique "thermostable chimeric DNA polymerases". For example, substitution of the Tma DNA polymerase coding sequence 15 comprising codons about 291 through about 484 for the Tag DNA polymerase I codons 289-422 would yield a novel thermostable DNA polymerase containing the 5' to 3' exonuclease domain of Taq DNA polymerase (1-289), the 3' to 5' exonuclease domain of Tma DNA polymerase 20 (291-484), and the DNA polymerase domain of Tag DNA polymerase (423-832). Alternatively, the 5' to exonuclease domain and the 3' to 5' exonuclease domains of Tma DNA polymerase (ca. codons 1-484) may be fused to the DNA polymerase (dNTP binding and primer/template 25 binding domains) portions of Tag DNA polymerase (ca. codons 423-832).

As is apparent, the donors and recipients for the creation of "thermostable chimeric DNA polymerase" by "domain shuffling" need not be limited to Tag and Tma Other thermostable polymerases 30 DNA polymerases. Tag and Tma DNA analogous domains as provide Furthermore, the 5' to 3' exonuclease polymerases. domain may derive from a thermostable DNA polymerase with altered 5' to 3' nuclease activity. For example, 35 the 1 to 289 5' to 3' nuclease domain of Tag DNA polymerase may derive from a Gly (46) to Asp mutant

form of the <u>Tag</u> polymerase gene. Similarly, the 5' to 3' nuclease and 3' to 5' nuclease domains of <u>Tma</u> DNA polymerase may encode a 5' to 3' exonuclease deficient domain, and be retrieved as a <u>Tma</u> Gly (37) to Asp amino acid 1 to 484 encoding DNA fragment or alternatively a truncated Met 140 to amino acid 484 encoding DNA fragment.

While any of a variety of means may be used to generate chimeric DNA polymerase coding sequences 10 (possessing novel properties), a preferred method employs "overlap" PCR. In this method, the intended junction sequence is designed into the PCR primers (at their 5'-ends). Following the initial amplification of the individual domains, the various products are diluted (ca. 100 to 1000-fold) and combined, denatured, annealed, extended, and then the final forward and reverse primers are added for an otherwise standard PCR.

Those of skill in the art recognize that the above thermostable DNA polymerases with attenuated 5' to 3' 20 exonuclease activity are most easily constructed by recombinant DNA techniques. When one desires to produce one of the mutant enzymes of the present 5′ 3′ exonuclease invention, with attenuated to activity or a derivative or homologue of those enzymes, 25 the production of a recombinant form of the enzyme typically involves the construction of an expression vector, the transformation of a host cell with the vector, and culture of the transformed host cell under conditions such that expression will occur.

To construct the expression vector, a DNA is obtained that encodes the mature (used here to include all chimeras or muteins) enzyme or a fusion of the mutant polymerase to an additional sequence that does not destroy activity or to an additional sequence to cleavable under controlled conditions (such as treatment with peptidase) to give an active protein.

The coding sequence is then placed in operable linkage with suitable control sequences in an expression vector. The vector can be designed to replicate autonomously in the host cell or to integrate into the 5 chromosomal DNA of the host cell. The vector is used to transform a suitable host, and the transformed host is cultured under conditions suitable for expression of the recombinant polymerase.

Each of the foregoing steps can be done in a For example, the desired coding 10 variety of ways. sequence may be obtained from genomic fragments and used directly in appropriate hosts. The construction for expression vectors operable in a variety of hosts made using appropriate replicons and 15 sequences, as set forth generally below. Construction of suitable vectors containing the desired coding and standard ligation employs sequences control restriction techniques that are well understood in the Isolated plasmids, DNA sequences, or synthesized 20 oligonucleotides are cleaved, modified, and religated in the form desired. Suitable restriction sites can, if not normally available, be added to the ends of the coding sequence so as to facilitate construction of an expression vector, as exemplified below.

Site-specific DNA cleavage is performed by treating 25 with suitable restriction enzyme (or enzymes) under conditions that are generally understood in the art and commercially the manufacturers of by specifi**e**d See, e.g., New England available restriction enzymes. In general, about 1 µg of 30 Biolabs, Product Catalog. plasmid or other DNA is cleaved by one unit of enzyme in about 20  $\mu$ l of buffer solution; in the examples below, an excess of restriction enzyme is generally ensure complete digestion of to 35 Incubation times of about one to two hours at about are typical, although variations can 37°C

tolerated. After each incubation, protein is removed extraction with phenol and chloroform; extraction can be followed by ether extraction and recovery of the DNA from aqueous fractions 5 precipitation with ethanol. If desired, separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. See, e.g., Methods in Enzymology, 1980, <u>65</u>:499-560.

Restriction-cleaved fragments with single-strand "overhanging" termini can be made blunt-ended (double-strand ends) by treating with the large fragment of E. coli DNA polymerase I (Klenow) in the presence of the four deoxynucleoside triphosphates 15 (dNTPs) using incubation times of about 15 to minutes at 20°C to 25°C in 50 mM Tris-Cl pH 7.6, 50 mM NaCl, 10 mM MgCl2, 10 mM DTT, and 5 to 10 µM dNTPs. The Klenow fragment fills in at 5' protruding ends, but chews back protruding 3' single strands, even though 20 the four dNTPs are present. If desired, selective repair can be performed by supplying only one of the, or selected, dNTPs within the limitations dictated by the nature of the protruding ends. After treatment with Klenow, the mixture is extracted with 25 phenol/chloroform and ethanol precipitated. Similar results can be achieved using S1 nuclease, because treatment under appropriate conditions with S1 nuclease results in hydrolysis of any single-stranded portion of a nucleic acid.

30 Synthetic oligonucleotides can be prepared using the triester method of Matteucci et al., 1981, J. Am. Chem. Soc. 103:3185-3191, or automated synthesis methods. Kinasing of single strands prior to annealing or for labeling is achieved using an excess, e.g., 35 approximately 10 units, of polynucleotide kinase to 0.5 μM substrate in the presence of 50 mM Tris, pH 7.6,

10 mM MgCl $_2$ , 5 mM dithiothreitol (DTT), and 1 to 2  $\mu$ M ATP. If kinasing is for labeling of probe, the ATP will contain high specific activity  $\gamma$ - $^{32}$ P.

Ligations are performed in 15-30 µl volumes under 5 the following standard conditions and temperatures: 20 mM Tris-Cl, pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 33 μg/ml BSA, 10 mM-50 mM NaCl, and either 40 µM ATP and 0.01-0.02 (Weiss) units T4 DNA ligase at 0°C (for with complementary of fragments ligation 10 single-stranded ends) or 1 mM ATP and 0.3-0.6 units T4 "blunt end" 14°C (for ligase at ligations of fragments with Intermolecular complementary ends are usually performed at 33-100  $\mu$ g/ml total DNA concentrations (5 to 100 nM total ends Intermolecular blunt end ligations 15 concentration). (usually employing a 20 to 30 fold molar excess of linkers, optionally) are performed at 1 µM total ends concentration.

In vector construction, the vector fragment is commonly treated with bacterial or calf intestinal alkaline phosphatase (BAP or CIAP) to remove the 5' phosphate and prevent religation and reconstruction of the vector. BAP and CIAP digestion conditions are well known in the art, and published protocols usually accompany the commercially available BAP and CIAP enzymes. To recover the nucleic acid fragments, the preparation is extracted with phenol-chloroform and ethanol precipitated to remove the phosphatase and purify the DNA. Alternatively, religation of unwanted vector fragments can be prevented by restriction enzyme digestion before or after ligation, if appropriate restriction sites are available.

For portions of vectors or coding sequences that require sequence modifications, a variety of 35 site-specific primer-directed mutagenesis methods are available. The polymerase chain reaction (PCR) can be

WO 92/06200

used to perform site-specific mutagenesis. In another technique now standard in the art, a synthetic oligonucleotide encoding the desired mutation is used as a primer to direct synthesis of a complementary 5 nucleic acid sequence of a single-stranded vector, such as pBS13+, that serves as a template for construction of the extension product of the mutagenizing primer. DNA is transformed into a host The mutagenized bacterium, and cultures of the transformed bacteria are 10 plated and identified. The identification of modified vectors may involve transfer of the DNA of selected transformants to a nitrocellulose filter or other membrane and the "lifts" hybridized with kinased synthetic primer at a temperature that permits 15 hybridization of an exact match to the modified sequence but prevents hybridization with the original strand. Transformants that contain DNA that hybridizes with the probe are then cultured and serve as a reservoir of the modified DNA.

In the constructions set forth below, correct 20 ligations for plasmid construction are confirmed by first transforming E. coli strain DG101 or another suitable host with the ligation mixture. Successful transformants are selected by ampicillin, tetracycline 25 or other antibiotic resistance or sensitivity or by using other markers, depending on the mode of plasmid construction, as is understood in the art. Plasmids from the transformants are then prepared according to the method of Clewell et al., 1969, Proc. Natl. Acad. 30 Sci. USA 62:1159, optionally following chloramphenicol amplification (Clewell, 1972, J. Bacteriol. 110:667). Another method for obtaining plasmid DNA is described as the "Base-Acid" extraction method at page 11 of the Bethesda Research Laboratories publication Focus, 35 volume 5, number 2, and very pure plasmid DNA can be obtained by replacing steps 12 through 17 of the protocol with CsCl/ethidium bromide ultracentrifugation of the DNA. The isolated DNA is analyzed by restriction enzyme digestion and/or sequenced by the dideoxy method of Sanger et al., 1977, Proc. Natl. 5 Acad. Sci. USA 74:5463, as further described by Messing et al., 1981, Nuc. Acids Res. 9:309, or by the method of Maxam et al., 1980, Methods in Enzymology 65:499.

The control sequences, expression vectors, and transformation methods are dependent on the type of 10 host cell used to express the gene. Generally, procaryotic, yeast, insect, or mammalian cells are used as hosts. Procaryotic hosts are in general the most efficient and convenient for the production of recombinant proteins and are therefore preferred for the expression of the thermostable DNA polymerases of the present invention.

The procaryote most frequently used to express recombinant proteins is E. coli. For cloning and sequencing, and for expression of constructions under 20 control of most bacterial promoters, E. coli K12 strain MM294, obtained from the E. coli Genetic Stock Center under GCSC #6135, can be used as the host. expression vectors with the  $P_L N_{RBS}$  control sequence,  $\underline{E}$ . coli K12 strain MC1000 lambda lysogen, N7N53CI857 25 SusP<sub>80</sub>, ATCC 39531, may be used. E. coli DG116, which was deposited with the ATCC (ATCC 53606) on April 7, 1987, and E. coli KB2, which was deposited with the ATCC (ATCC 53075) on March 29, 1985, are also useful For M13 phage recombinants,  $\underline{\mathbf{E}}$ . host cells. 30 strains susceptible to phage infection, such as E. coli K12 strain DG98, are employed. The DG98 strain was deposited with the ATCC (ATCC 39768) on July 13, 1984.

However, microbial strains other than <u>E. coli</u> can also be used, such as bacilli, for example <u>Bacillus</u>

35 <u>subtilis</u>, various species of <u>Pseudomonas</u>, and other bacterial strains, for recombinant expression of the

thermostable DNA polymerases of the present invention. In such procaryotic systems, plasmid vectors that contain replication sites and control sequences derived from the host or a species compatible with the host are typically used.

For example, E. coli is typically transformed using derivatives of pBR322, described by Bolivar et al., 1977, <u>Gene</u> <u>2</u>:95. Plasmid pBR322 contains genes for ampicillin and tetracycline resistance. These drug 10 resistance markers can be either retained or destroyed in constructing the desired vector and so help to detect the presence of a desired recombinant. Commonly used procaryotic control sequences, i.e., a promoter transcription initiation, optionally with 15 operator, along with a ribosome binding site sequence, include the B-lactamase (penicillinase) and lactose (lac) promoter systems (Chang et al., 1977, Nature the tryptophan <u>198</u>:1056), (trp) promoter (Goeddel et al., 1980, Nuc. Acids Res. 8:4057), and the 20 lambda-derived P<sub>I</sub>, promoter (Shimatake et al., 1981, Nature 292:128) and N-gene ribosome binding site  $(N_{
m RBS})$ . A portable control system cassette is set forth in United States Patent No. 4,711,845, issued December 8, 1987. This cassette comprises a  $P_{T}$ 25 promoter operably linked to the  $N_{\mbox{RBS}}$  in turn positioned upstream of a third DNA sequence having at least one restriction site that permits cleavage within six bp 3' of the  $N_{\mbox{RBS}}$  sequence. Also useful is the phosphatase A (phoA) system described by Chang et al. in European 30 Patent Publication No. 196,864, published October 8, 1986. However, available any promoter compatible with procaryotes can be used to construct a modified thermostable DNA polymerase expression vector of the invention.

In addition to bacteria, eucaryotic microbes, such as yeast, can also be used as recombinant host cells.

Laboratory strains of Saccharomyces cerevisiae, Baker's yeast, are most often used, although a number of other available. While strains are commonly employing the two micron origin of replication are Enz. 101:307), 5 common (Broach, 1983, Meth. plasmid vectors suitable for yeast expression are known (see, for example, Stinchcomb et al., 1979, Nature 282:39; Tschempe et al., 1980, Gene 10:157; and Clarke et al., 1983, Meth. Enz. 101:300). Control sequences 10 for yeast vectors include promoters for the synthesis of glycolytic enzymes (Hess et al., 1968, J. Enzyme Req. 7:149; Holland et al., 1978, Biotechnology 17:4900; and Holland et al., 1981, J. Biol. Chem. Additional promoters known in the 256:1385). 15 include the promoter for 3-phosphoglycerate kinase (Hitzeman et al., 1980, J. Biol. Chem. 255:2073) and glycolytic enzymes, such as for other those glyceraldehyde 3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-3-phosphoglycerate mutase, isomerase, 20 phosphate triosephosphate isomerase, pyruvate kinase, glucokinase. Other isomerase, and phosphoglucose that have the additional advantage transcription controlled by growth conditions are the dehydrogenase for alcohol 25 promoter regions isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and maltose and galactose utilization responsible for (Holland, supra).

Terminator sequences may also be used to enhance expression when placed at the 3' end of the coding sequence. Such terminators are found in the 3' untranslated region following the coding sequences in yeast-derived genes. Any vector containing a yeast-compatible promoter, origin of replication, and

other control sequences is suitable for use in constructing yeast expression vectors for the thermostable DNA polymerases of the present invention.

The nucleotide sequences which code 5 thermostable DNA polymerases of the present invention can also be expressed in eucaryotic host cell cultures derived from multicellular organisms. See. example, <u>Tissue</u> <u>Culture</u>, Academic Press, Cruz and Patterson, editors (1973). Useful host cell lines 10 include COS-7, COS-A2, CV-1, murine cells such as murine myelomas N51 and VERO, HeLa cells, and Chinese hamster ovary (CHO) cells. Expression vectors for such cells ordinarily include promoters and sequences compatible with mammalian cells such as, for 15 example, the commonly used early and late promoters from Simian Virus 40 (SV 40) (Fiers et al., 1978, Nature 273:113), or other viral promoters such as those derived from polyoma, adenovirus 2, bovine papilloma virus (BPV), oravian sarcoma viruses, 20 immunoglobulin promoters and heat shock promoters. A system for expressing DNA in mammalian systems using a BPV vector system is disclosed in U.S. Patent No. 4,419,446. A modification of this system is described in U.S. Patent No. 4,601,978. General aspects of 25 mammalian cell host system transformations have been described by Axel, U.S. Patent No. 4,399,216. "Enhancer" regions are also important in optimizing expression; these are, generally, sequences found upstream of the promoter region. Origins 30 replication may be obtained, if needed, from viral sources. However, integration into the chromosome is a common mechanism for DNA replication in eucaryotes.

Plant cells can also be used as hosts, and control sequences compatible with plant cells, such as the 35 nopaline synthase promoter and polyadenylation signal sequences (Depicker et al., 1982, J. Mol. Appl. Gen.

1:561) are available. Expression systems employing insect cells utilizing the control systems provided by baculovirus vectors have also been described (Miller et al., 1986, Genetic Engineering (Setlow et al., eds.,

- 5 Plenum Publishing) <u>8</u>:277-297). Insect cell-based expression can be accomplished in <u>Spodoptera frugipeida</u>. These systems can also be used to produce recombinant thermostable polymerases of the present invention.
- Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described by Cohen, 1972, <a href="Proc. Natl.Acad.Sci.USA 69:2110">Proc. Natl.Acad.Sci.USA 69:2110</a> is used for procaryotes or other
- 15 cells that contain substantial cell wall barriers. Infection with <u>Agrobacterium tumefaciens</u> (Shaw <u>et al.</u>, 1983, <u>Gene 23</u>:315) is used for certain plant cells. For mammalian cells, the calcium phosphate precipitation method of Graham and van der Eb, 1978,
- 20 <u>Virology</u> <u>52</u>:546 is preferred. Transformations into yeast are carried out according to the method of Van Solingen <u>et al.</u>, 1977, <u>J. Bact.</u> <u>130</u>:946 and Hsiao <u>et al.</u>, 1979, <u>Proc. Natl. Acad. Sci. USA</u> <u>76</u>:3829.
- Once the desired thermostable DNA polymerase with 25 altered 5' to 3' exonuclease activity has been expressed in a recombinant host cell, purification of the protein may be desired. Although a variety of purification procedures can be used to purify the recombinant thermostable polymerases of the invention,
- 30 fewer steps may be necessary to yield an enzyme preparation of equal purity. Because  $\underline{E}$ .  $\underline{coli}$  host proteins are heat-sensitive, the recombinant thermostable DNA polymerases of the invention can be substantially enriched by heat inactivating the crude
- 35 lysate. This step is done in the presence of a sufficient amount of salt (typically 0.2-0.3 M ammonium

sulfate) to ensure dissociation of the thermostable DNA polymerase from the host DNA and to reduce ionic interactions of thermostable DNA polymerase with other cell lysate proteins.

In addition, the presence of 0.3 M ammonium sulfate promotes hydrophobic interaction with sepharose column. Hydrophobic interaction chromatography is a separation technique in which substances are separated on the basis of differing 10 strengths of hydrophobic interaction with an uncharged bed material containing hydrophobic groups. Typically, the column is first equilibrated under conditions favorable to hydrophobic binding, such as high ionic A descending salt gradient may then be used strength. 15 to elute the sample.

According to the invention, an aqueous mixture (containing the recombinant thermostable DNA polymerase with altered 5' to 3' exonuclease activity) is loaded containing column a relatively 20 hydrophobic gel such as phenyl sepharose (manufactured by Pharmacia) or Phenyl TSK (manufactured by Toyo Soda). To promote hydrophobic interaction with a phenyl sepharose column, a solvent is used contains, for example, greater than or equal to 0.3 M 25 ammonium sulfate, with 0.3 M being preferred, or greater than or equal to 0.5 M NaCl. The column and the sample are adjusted to 0.3 M ammonium sulfate in 50 mM Tris (pH 7.5) and 1.0 mM EDTA ("TE") buffer that also contains 0.5 mM DTT, and the sample is applied to 30 the column. The column is washed with the 0.3 M ammonium sulfate buffer. The enzyme may then be eluted with solvents that attenuate hydrophobic interactions, as decreasing salt gradients, ethylene propylene glycol, or urea.

35 For long-term stability, the thermostable DNA polymerase enzymes of the present invention can be

stored in a buffer that contains one or more non-ionic Such detergents are generally polymeric detergents. those that have a molecular weight in the range of approximately 100 to 250,000 daltons, preferably about 5 4,000 to 200,000 daltons, and stabilize the enzyme at a pH of from about 3.5 to about 9.5, preferably from Examples of such detergents include about 4 to 8.5. pages 295-298 of McCutcheon's specified on those North American edition Emulsifiers & Detergents, 10 (1983), published by the McCutcheon Division of MC Publishing Co., 175 Rock Road, Glen Rock, NJ (USA) and copending Serial No. 387,003, filed July 28, 1989, each of which is incorporated herein by reference.

Preferably, the detergents are selected from the 15 group comprising ethoxylated fatty alcohol ethers and lauryl ethers, ethoxylated alkyl phenols, octylphenoxy compounds, modified oxyethylated polyethoxy ethanol straight-chain oxypropylated and/or polyethylene glycol monooleate compounds, polysorbate 20 compounds, and phenolic fatty alcohol ethers. particularly preferred are Tween 20, a polyoxyethylated sorbitan monolaurate from ICI Americas Inc., Wilmington, DE, and Iconol NP-40, an ethoxylated alkyl phenol (nonyl) from BASF Wyandotte Corp., Parsippany, 25 NJ.

The thermostable enzymes of this invention may be used for any purpose in which such enzyme activity is ecessary or desired.

DNA sequencing by the Sanger dideoxynucleotide

30 method (Sanger et al., 1977, Proc. Natl. Acad. Sci. USA

74:5463-5467) has undergone significant refinement in
recent years, including the development of novel
vectors (Yanisch-Perron et al., 1985, Gene 33:103-119),
base analogs (Mills et al., 1979, Proc. Natl. Acad.

35 Sci. USA 76:2232-2235, and Barr et al., 1986,
BioTechniques 4:428-432), enzymes (Tabor et al., 1987,

Proc. Natl. Acad. Sci. USA 84:4763-4771, and Innis, et al., 1988, Proc. Natl. Acad. Sci. 85:9436:9440), and instruments for partial automation of DNA sequence analysis (Smith et al., 1986, Nature 5 321:674-679; Prober et al., 1987, Science 238:336-341; Ansorge et <u>al.,</u> 1987, Nuc. Acids Res. <u>15</u>:4593-4602). The basic dideoxy sequencing procedure involves (i) annealing an oligonucleotide primer to a suitable single or denatured double stranded 10 template; (ii) extending the primer with DNA polymerase four separate reactions, each containing α-labeled dNTP or ddNTP (alternatively, a labeled primer can be used), a mixture of unlabeled dNTPs, and one chain-terminating dideoxynucleotide-5'-triphosphate 15 (ddNTP); (iii) resolving the four sets of reaction products on a high-resolution polyacrylamide-urea gel; and (iv) producing an autoradiographic image of the gel that can be examined to infer the DNA sequence. Alternatively, fluorescently labeled primers 20 nucleotides can be used to identify the reaction products. Known dideoxy sequencing methods utilize a DNA polymerase such as the Klenow fragment of E. coli polymerase I, reverse transcriptase, Tag DNA polymerase, or a modified T7 DNA polymerase.

25 The introduction of commercial kits has vastly simplified the art, making DNA sequencing a routine technique for any laboratory. However, there is still a need in the art for sequencing protocols that work with nucleic acids that contain secondary 30 structure such as palindromic hairpin loops and with G+C-rich DNA. Single stranded DNAs can form secondary structure, such as a hairpin loop, that can seriously interfere with a dideoxy sequencing protocol, through improper termination in the extension reaction, 35 or in the case of an enzyme with 5' to 3' exonuclease activity, cleavage of the template strand at the

Since high temperature juncture of the hairpin. structure, ability secondary the destabilizes conduct the extension reaction at a high temperature, i.e., 70-75°C, with a thermostable DNA polymerase 5 results in a significant improvement in the sequencing secondary contains such of DNA that with temperatures compatible polymerase However, extension do not eliminate all secondary structure. exonuclease-deficient DNA thermostable 10 polymerase would be a further improvement in the art, since the polymerase could synthesize through hairpin in a strand displacement reaction, rather than cleaving the template, resulting in an improper termination, i.e., an extension run-off fragment.

15 As an alternative to basic dideoxy sequencing, cycle dideoxy sequencing is a linear, asymmetric amplification of target sequences in the presence of dideoxy chain terminators. A single cycle produces a family of extension products of all possible lengths.

- 20 Following denaturation of the extension reaction product from the DNA template, multiple cycles of primer annealing and primer extension occur in the presence of dideoxy terminators. The process is distinct from PCR in that only one primer is used, the
- 25 growth of the sequencing reaction products in each cycle is linear, and the amplification products are heterogeneous in length and do not serve as template for the next reaction. Cycle dideoxy sequencing is a technique providing advantages for laboratories using
- 30 automated DNA sequencing instruments and for other high volume sequencing laboratories. It is possible to directly sequence genomic DNA, without cloning, due to the specificity of the technique and the increased amount of signal generated. Cycle sequencing protocols
- 35 accommodate single and double stranded templates, including genomic, cloned, and PCR-amplified templates.

Thermostable DNA polymerases have several advantages in cycle sequencing: they tolerate the stringent annealing temperatures which are required for specific hybridization of primer to genomic targets as tolerating the multiple cycles as of high temperature denaturation which occur in each cycle. Performing the extension reaction at high temperatures, i.e., 70-75°C, results in a significant improvement in sequencing results with DNA that contains secondary 10 structure. due to the destabilization of secondary structure. However, such temperatures will eliminate all secondary structure. Α to 3' exonuclease-deficient thermostable DNA polymerase would further improvement in the art, since the 15 polymerase could synthesize through the hairpin in a strand displacement reaction, rather than cleaving the template and creating an improper termination. Additionally, like PCR, cycle sequencing suffers from the phenomenon of product strand renaturation. 20 case of a thermostable DNA polymerase possessing 5' to 3' exonuclease activity, extension of a primer into a double stranded region created by product strand renaturation will result in cleavage of the renatured complementary product strand. The cleaved strand will 25 be shorter and thus appear as an improper termination. addition, the correct. previously synthesized termination signal will be attenuated. A thermostable polymerase deficient in 5' to 3' exonuclease activity will improve the art, in that such extension 30 product fragments will not be formed. A variation of cycle sequencing, involves the simultaneous generation sequencing ladders for each strand of a double stranded template while sustaining some degree amplification (Ruano and Kidd, Proc. Natl. Acad. Sci. 35 USA, 1991 <u>88</u>:2815-2819). This method of amplification and sequencing would benefit in a similar

fashion as stranded cycle sequencing from the use of a thermostable DNA polymerase deficient in 5' to 3' exonuclease activity.

In a particularly preferred embodiment, the enzymes 5 in which the 5' to 3' exonuclease activity has been eliminated catalyze the nucleic reduced or amplification reaction known as PCR, and as stated above, with the resultant effect of producing a better yield of desired product than is achieved with the 10 respective native enzymes which have greater amounts of the 5' to 3' exonuclease activity. Improved yields are the result of the inability to degrade previously synthesized product caused by 5' to 3' exonuclease This process for amplifying nucleic acid 15 sequences is disclosed and claimed in U.S. Patent Nos. 4,683,202 and 4,865,188, each of which is incorporated PCR nucleic reference. The herein bv amplification method involves amplifying at least one specific nucleic acid sequence contained in a nucleic 20 acid or a mixture of nucleic acids and in the most common embodiment, produces double-stranded DNA. Aside from improved yields, thermostable DNA polymerases with attenuated 5' to 3' exonuclease activity exhibit an improved ability to generate longer PCR products, an 25 improved ability to produce products from G+C-rich templates and an improved ability to generate PCR products and DNA sequencing ladders from templates with a high degree of secondary structure.

For ease of discussion, the protocol set forth assumes that the specific sequence to 30-below amplified is contained in a double-stranded nucleic However, the process is equally useful acid. amplifying single-stranded nucleic acid, such as mRNA, although in the preferred embodiment the ultimate DNA. Tn still double-stranded 35 product is amplification of a single-stranded nucleic acid,

first step involves the synthesis of a complementary strand (one of the two amplification primers can be used for this purpose), and the succeeding steps proceed as in the double-stranded amplification process 5 described below.

This amplification process comprises the steps of:

- (a) contacting each nucleic acid strand with four 10 different nucleoside triphosphates oligonucleotide primers for each specific sequence being amplified, wherein each primer is selected to be substantially complementary to the different strands of the specific sequence, such that the extension product 15 synthesized from one primer, when separated from its complement, can serve as a template for synthesis of the extension product of the other primer. contacting being at a temperature that hybridization of each primer to a complementary nucleic 20 acid strand:
- (b) contacting each nucleic acid strand, at the same time as or after step (a), with a thermostable DNA polymerase of the present invention that enables combination of the nucleoside triphosphates to form 25 primer extension products complementary to each strand of the specific nucleic acid sequence;
- (c) maintaining the mixture from step (b) at an effective temperature for an effective time to promote the activity of the enzyme and to synthesize, for each 30 different sequence being amplified, an extension product of each primer that is complementary to each nucleic acid strand template, but not so high as to separate each extension product from the complementary strand template;
- 35 (d) heating the mixture from step (c) for an effective time and at an effective temperature to

separate the primer extension products from the templates on which they were synthesized to produce single-stranded molecules but not so high as to denature irreversibly the enzyme;

- (e) cooling the mixture from step (d) for an effective time and to an effective temperature to promote hybridization of a primer to each of the single-stranded molecules produced in step (d); and
- (f) maintaining the mixture from step (e) at an 10 effective temperature for an effective time to promote the activity of the enzyme and to synthesize, for each amplified, an extension different sequence being product of each primer that is complementary to each nucleic acid template produced in step (d) but not so 15 high as to separate each extension product from the complementary strand template. The effective times and temperatures in steps (e) and (f) may coincide, so that steps (e) and (f) can be carried out simultaneously. Steps (d)-(f) are repeated until the desired level of 20 amplification is obtained.

The amplification method is useful not only for producing large amounts of a specific nucleic acid sequence of known sequence but also for producing nucleic acid sequences that are known to exist but are 25 not completely specified. One need know only a sufficient number of bases at both ends of the sequence in sufficient detail so that two oligonucleotide primers can be prepared that will hybridize different strands of the desired sequence at relative 30 positions along the sequence such that an extension product synthesized from one primer, when separated from the template (complement), can serve as a template for extension of the other primer into a nucleic acid sequence of defined length. The greater the knowledge 35 about the bases at both ends of the sequence, the greater can be the specificity of the primers for the

target nucleic acid sequence and the efficiency of the process and specificity of the reaction.

In any case, an initial copy of the sequence to be amplified must be available, although the sequence need 5 not be pure or a discrete molecule. In general, the amplification process involves a chain reaction for producing, in exponential quantities relative to the number of reaction steps involved, at least specific nucleic acid sequence given that (a) the ends 10 of the required sequence are known in sufficient detail that oligonucleotides can be synthesized that will hybridize to them and (b) that a small amount of the sequence is available to initiate the chain reaction. The product of the chain reaction will be a discrete 15 nucleic acid duplex with termini corresponding to the 5' ends of the specific primers employed.

nucleic Any acid sequence, in purified nonpurified form, can be utilized as the starting nucleic acid(s), provided it contains or is suspected 20 to contain the specific nucleic acid sequence one desires to amplify. The nucleic acid to be amplified can be obtained from any source, for example, from plasmids such as pBR322, from cloned DNA or RNA, or from natural DNA or RNA from any source, including 25 bacteria, yeast, viruses, organelles, and organisms such as plants and animals. DNA or RNA may be extracted from blood, tissue material such chorionic villi, or amniotic cells by a variety of techniques. See, e.g., Maniatis <u>et</u> <u>al.,</u> 30 Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) pp. 280-281. Thus, the process may employ, example, DNA or RNA, including messenger RNA, which DNA or RNA may be single-stranded or double-stranded. 35 addition, a DNA-RNA hybrid that contains one strand of each may be utilized. A mixture of any of these

WO 92/06200 PCT/US91/07035

nucleic acids can also be employed as can nucleic acids produced from a previous amplification reaction (using the same or different primers). The specific nucleic acid sequence to be amplified can be only a fraction of a large molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire nucleic acid.

The sequence to be amplified need not be present initially in a pure form; the sequence can be a minor 10 fraction of a complex mixture, such as a portion of the whole DNA human gene contained in B-globin 1985, Science in Saiki al., exemplified <u>et</u> 230:1530-1534) or a portion of a nucleic acid sequence due to a particular microorganism, which organism might 15 constitute only a very minor fraction of a particular biological sample. The cells can be directly used in the amplification process after suspension in hypotonic buffer and heat treatment at about 90°C-100°C until cell lysis and dispersion of intracellular components 20 occur (generally 1 to 15 minutes). After the heating step, the amplification reagents may be added directly to the lysed cells. The starting nucleic acid sequence can contain more than one desired specific nucleic acid sequence. The amplification process is useful not only 25 for producing large amounts of one specific nucleic acid sequence but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules.

Primers play a key role in the PCR process. The 30 word "primer" as used in describing the amplification process can refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the fragment to be amplified or where one employs the 35 degenerate primer process described in PCT Application No. 91/05753, filed August 13, 1991. For instance, in

the case where a nucleic acid sequence is inferred from protein sequence information, a collection of primers containing sequences representing all possible codon variations based on degeneracy of the genetic code can 5 be used for each strand. One primer from this collection will be sufficiently homologous with a portion of the desired sequence to be amplified so as to be useful for amplification.

In addition, more than one specific nucleic acid 10 sequence can be amplified from the first nucleic acid or mixture of nucleic acids, so long as the appropriate different oligonucleotide number of primers For example, if two different specific utilized. nucleic acid sequences are to be produced, four primers 15 are utilized. Two of the primers are specific for one of the specific nucleic acid sequences, and the other two primers are specific for the second specific nucleic acid sequence. In this manner, each of the two different specific sequences can be produced 20 exponentially by the present process.

A sequence within a given sequence can be amplified after a given number of amplification cycles to obtain greater specificity in the reaction by adding, after at least one cycle of amplification, a set of primers that complementary to internal sequences sequences that are not on the ends) of the sequence to be amplified. Such primers can be added at any stage will provide a shorter amplified fragment. Alternatively, a longer fragment can be prepared by 30 using primers with non-complementary ends but having some overlap with the primers previously utilized in the amplification.

Primers also play a key role when the amplification process is used for <u>in vitro</u> mutagenesis. The product 35 of an amplification reaction where the primers employed are not exactly complementary to the original template

will contain the sequence of the primer rather than the template, so introducing an in vitro mutation. In further cycles, this mutation will be amplified with an undiminished efficiency because no further mispaired 5 priming is required. The process of making an altered DNA sequence as described above could be repeated on the altered DNA using different primers to induce further sequence changes. In this way, a series of mutated sequences can gradually be produced wherein 10 each new addition to the series differs from the last in a minor way, but from the original DNA source sequence in an increasingly major way.

Because the primer can contain as part of sequence a non-complementary sequence, provided that a 15 sufficient amount of the primer contains a sequence that is complementary to the strand to be amplified, many other advantages can be realized. For example, a nucleotide sequence that is not complementary to the template sequence (such as, e.g., a promoter, linker, 20 coding sequence, etc.) may be attached at the 5' end of one or both of the primers and so appended to the amplification process. of the extension primer is added, sufficient cycles are run to achieve the desired amount of new template containing 25 the non-complementary nucleotide insert. This allows quantities the production of large of fragments in a relatively short period of time (e.g., two hours or less) using a simple technique.

Oligonucleotide primers can be prepared using any 30 suitable method. such as, for example, phosphotriester and phosphodiester methods described above, or automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used starting materials and can be synthesized 35 described by Beaucage et al., 1981, Tetrahedron Letters for synthesizing 22:1859-1862. One method

PCT/US91/07035

oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066. One can also use a primer that has been isolated from a biological source (such as a restriction endonuclease digest).

-50-

No matter what primers are used, however, reaction mixture must contain a template for PCR to occur, because the specific nucleic acid sequence is produced by using a nucleic acid containing sequence as a template. The first step involves 10 contacting each nucleic acid strand with four different nucleoside triphosphates and two oligonucleotide primers for each specific nucleic acid sequence being amplified or detected. If the nucleic acids to be amplified or detected are DNA, then the nucleoside 15 triphosphates are usually dATP, dCTP, dGTP, and dTTP, although various nucleotide derivatives can also be used in the process. For example, when using PCR for the detection of a known sequence in a sample of unknown sequences, dTTP is often replaced by dUTP in 20 order to reduce contamination between samples as taught in PCT Application No. 91/05210 filed July 23, 1991, incorporated herein by reference.

The concentration of nucleoside triphosphates can vary widely. Typically, the concentration is 50 to 200 25 µM in each dNTP in the buffer for amplification, and MgCl<sub>2</sub> is present in the buffer in an amount of 1 to 3 to activate the polymerase and increase specificity of the reaction. However, dNTP concentrations of 1 to 20 µM may be preferred for some 30 applications, such as DNA sequencing or generating radiolabeled probes at high specific activity.

The nucleic acid strands of the target nucleic acid serve as templates for the synthesis of additional nucleic acid strands, which are extension products of 35 the primers. This synthesis can be performed using any suitable method, but generally occurs in a buffered

aqueous solution, preferably at a pH of 7 to 9, most preferably about 8. To facilitate synthesis, a molar excess of the two oligonucleotide primers is added to the buffer containing the template strands. 5 practical matter, the amount of primer added will generally be in molar excess over the amount of complementary strand (template) when the sequence to be amplified is contained in a mixture of complicated long-chain nucleic acid strands. A large molar excess 10 is preferred to improve the efficiency of the process. Accordingly, primer:template ratios of at least 1000:1 generally employed for cloned or higher are templates, and primer: template ratios of about 108:1 or higher are generally employed for amplification from 15 complex genomic samples.

The mixture of template, primers, and nucleoside triphosphates is then treated according to whether the nucleic acids being amplified or detected are doubleacids single-stranded. Ιf the nucleic 20 single-stranded, then no denaturation step need be employed prior to the first extension cycle, and the reaction mixture is held at a temperature that promotes hybridization of the primer to its complementary target Such temperature is generally (template) sequence. 25 from about 35°C to 65°C or more, preferably about 37°C to 60°C for an effective time, generally from a few seconds to five minutes, preferably from 30 seconds to A hybridization temperature of 35°C to one minute. 70°C may be used for 5' to 3' exonuclease mutant 30 thermostable DNA polymerases. Primers that are nucleotides or longer in length are used to increase specificity of primer hybridization. primers require lower hybridization temperatures.

The complement to the original single-stranded 35 nucleic acids can be synthesized by adding the thermostable DNA polymerase of the present invention in

the presence of the appropriate buffer, dNTPs, and one or more oligonucleotide primers. If an appropriate single primer is added, the primer extension product will be complementary to the single-stranded nucleic 5 acid and will be hybridized with the nucleic acid strand in a duplex of strands of equal or unequal length (depending on where the primer hybridizes to the template), which may then be separated into single strands as described above to produce two single, 10 separated, complementary strands. A second primer would then be added so that subsequent cycles of primer extension would occur using both the original single-stranded nucleic acid and the extension product of the first primer as templates. Alternatively, two 15 or more appropriate primers (one of which will prime synthesis using the extension product of the other primer as a template) can be added single-stranded nucleic acid and the reaction carried out.

If the nucleic acid contains two strands, as in the 20 case of amplification of a double-stranded target or second-cycle amplification of a single-stranded target, the strands of nucleic acid must be separated before the primers are hybridized. This strand separation can 25 be accomplished by any suitable denaturing method, including physical, chemical or enzymatic means. preferred physical method of separating the strands of the nucleic acid involves heating the nucleic acid until complete (>99%) denaturation occurs. 30 heat denaturation involves temperatures ranging from about 80°C to 105°C for times generally ranging from about a few seconds to minutes, depending on the composition and size of the nucleic acid. Preferably, the effective denaturing temperature is 90°C-100°C for 35 a few seconds to 1 minute. Strand separation may also

be induced by an enzyme from the class of enzymes known

as helicases or the enzyme RecA, which has helicase activity and in the presence of ATP is known to The reaction conditions suitable for denature DNA. separating the strands of nucleic acids with helicases Hoffmann-Berling, described Kuhn by 5 are and techniques CSH-Quantitative Biology 43:63, using RecA are reviewed in Radding, 1982, Ann. Rev. The denaturation produces Genetics 16:405-437. separated complementary strands of equal or unequal 10 length.

If the double-stranded nucleic acid is denatured by heat, the reaction mixture is allowed to cool to a temperature that promotes hybridization of each primer to the complementary target (template) sequence. 15 temperature is usually from about 35°C to 65°C or more, depending on reagents, preferably 37°C to 60°C. an temperature is maintained hybridization effective time, generally a few seconds to minutes, and preferably 10 seconds to 1 minute. In practical terms, 20 the temperature is simply lowered from about 95°C to as low as 37°C, and hybridization occurs at a temperature within this range.

nucleic acid is singlethe Whether double-stranded, the thermostable DNA polymerase of the 25 present invention can be added prior to or during the denaturation step or when the temperature is being promoting for range reduced or is in the to hybridization. Although the thermostability of polymerases of the invention allows one to add such 30 polymerases to the reaction mixture at any time, one can substantially inhibit non-specific amplification by adding the polymerase to the reaction mixture at a point in time when the mixture will not be cooled below temperature. hybridization stringent 35 hybridization, the reaction mixture is then heated to or maintained at a temperature at which the activity of

the enzyme is promoted or optimized, i.e., a temperature sufficient to increase the activity of the in facilitating synthesis of the extension products from the hybridized primer 5 template. The temperature must actually be sufficient to synthesize an extension product of each primer that is complementary to each nucleic acid template, but must not be so high as to denature each extension product from its complementary template (i.e., 10 temperature is generally less than about 80°C to 90°C).

Depending on the nucleic acid(s) employed, the typical temperature effective for this synthesis reaction generally ranges from about 40°C to 80°C, preferably 50°C to 75°C. The temperature 15 preferably ranges from about 65°C to 75°C for the thermostable DNA polymerases of the present invention. The period of time required for this synthesis may range from about 10 seconds to several minutes or more, depending mainly on the temperature, the length of the 20 nucleic acid, the enzyme, and the complexity of the nucleic acid mixture. The extension time is usually about 30 seconds to a few minutes. If the nucleic acid is longer, a longer time period is generally required for complementary strand synthesis.

25 The newly synthesized strand and the complement nucleic acid strand form a double-stranded molecule is that used in the succeeding steps amplification process. In the next step, the strands of the double-stranded molecule are separated by heat 30 denaturation at a temperature and for a time effective to denature the molecule, but not at a temperature and for a period so long that the thermostable enzyme is completely and irreversibly denatured or inactivated. After this denaturation of template, the temperature is 35 decreased to a level that promotes hybridization of the

primer to the complementary single-stranded molecule (template) produced from the previous step, as described above.

After this hybridization step, or concurrently with 5 the hybridization step, the temperature is adjusted to a temperature that is effective to promote the activity of the thermostable enzyme to enable synthesis of a primer extension product using as a template both the newly synthesized and the original strands. The 10 temperature again must not be so high as to separate (denature) the extension product from its template, as described above. Hybridization may occur during this step, so that the previous step of cooling after denaturation is not required. In such a case, using 15 simultaneous steps, the preferred temperature range is 50°C to 70°C.

The heating and cooling steps involved in one cycle of strand separation, hybridization, and extension product synthesis can be repeated as many times as 20 needed to produce the desired quantity of the specific The only limitation is the nucleic acid sequence. thermostable enzyme, amount of the primers, Usually, from 15 to nucleoside triphosphates present. 30 cycles are completed. For diagnostic detection of 25 amplified DNA, the number of cycles will depend on the nature of the sample, the initial target concentration in the sample and the sensitivity of the detection amplification. For after used process sensitivity of detection, fewer cycles will be required 30 if the sample being amplified is pure and the initial If the sample is a target concentration is high. complex mixture of nucleic acids and the initial target concentration is low, more cycles will be required to amplify the signal sufficiently for detection. 35 general amplification and detection, the process is repeated about 15 times. When amplification is used to

generate sequences to be detected with labeled sequence-specific probes and when human genomic DNA is the target of amplification, the process is repeated 15 to 30 times to amplify the sequence sufficiently so 5 that a clearly detectable signal is produced, i.e., so that background noise does not interfere with detection.

No additional nucleotides, primers, or thermostable enzyme need be added after the initial provided that no key reagent has been exhausted and 10 that the enzyme has not become denatured irreversibly inactivated, in which case additional polymerase or other reagent would have to be added for the reaction to continue. After the appropriate number of cycles has been completed to produce the desired 15 amount of the specific nucleic acid sequence, the reaction can be halted in the usual manner, e.g., by inactivating the enzyme by adding EDTA, phenol, SDS, or CHCl3 or by separating the components of the reaction.

The amplification process can be conducted 20 continuously. In one embodiment of an automated process, the reaction mixture can be temperature cycled that temperature is the programmed controlled at a certain level for a certain time. such instrument for this purpose is the automated 25 machine handling for the amplification developed marketed and by. Perkin-Elmer Cetus Instruments. Detailed instructions for carrying out PCR with the instrument are available upon purchase of the instrument.

30 The thermostable DNA polymerases of the present invention with altered 5' to 3' exonuclease activity are very useful in the diverse processes in which amplification of a nucleic acid sequence by PCR is useful. The amplification method may be utilized to 35 clone a particular nucleic acid sequence for insertion into a suitable expression vector, as described in U.S.

pp. 11-19.

The vector may be used to Patent No. 4,800,159. transform an appropriate host organism to produce the gene product of the sequence by standard methods of recombinant DNA technology. Such cloning may involve 5 direct ligation into a vector using blunt-end ligation, or use of restriction enzymes to cleave at sites contained within the primers. Other processes suitable for the thermostable DNA polymerases of the present invention include those described in U.S. Patent Nos. 10 4,683,195 and 4,683,202 and European Patent Publication Nos. 229,701; 237,362; and 258,017; these patents and publications are incorporated herein by reference. addition, the present enzyme is useful in asymmetric PCR (see Gyllensten and Erlich, 1988, Proc. Natl. Acad. incorporated 15 Sci. USA <u>85</u>:7652-7656, herein reference); inverse PCR (Ochman et al., 1988, Genetics 120:621, incorporated herein by reference); and for DNA sequencing (see Innis et al., 1988, Proc. Natl. Acad. Sci. USA 85:9436-9440, and McConlogue et al., 1988, 20 Nuc. Acids Res. 16(20):9869), random amplification of cDNA ends (RACE), random priming PCR which is used to amplify a series of DNA fragments, and PCR processes with single sided specificity such as anchor PCR and ligation-mediated anchor PCR as described by Loh, E. in 25 METHODS: A Companion to Methods in Enzymology (1991) 2:

An additional process in which a 5' to 3' exonuclease deficient thermostable DNA polymerase would be useful is a process referred to as polymerase ligase 30 chain reaction (PLCR). As its name suggests, this process combines features of PCR with features of ligase chain reaction (LCR).

PLCR was developed in part as a technique to increase the specificity of allele-specific PCR in 35 which the low concentrations of dNTPs utilized (~1  $\mu$ M) limited the extent of amplification. In PLCR, DNA is

denatured and four complementary, but not adjacent, oligonucleotide primers are added with dNTPs, a thermostable DNA polymerase and a thermostable ligase.

The primers anneal to target DNA in a non-adjacent 5 fashion and the thermostable DNA polymerase causes the addition of appropriate dNTPs to the 3' end of the downstream primer to fill the gap between the non-adjacent primers and thus render the primers adjacent. The thermostable ligase will then ligate the 10 two adjacent oligonucleotide primers.

However, the presence of 5' to 3' exonuclease activity in the thermostable DNA polymerase significantly decreases the probability of closing the gap between the two primers because such activity 15 causes the excision of nucleotides small oligonucleotides from the 5' end of the downstream primer thus preventing ligation of the primers. Therefore, thermostable a DNA polymerase attenuated or eliminated 5' to 3' exonuclease activity 20 would be particularly useful in PLCR.

Briefly, the thermostable DNA polymerases of the present invention which have been mutated to have reduced, attenuated or eliminated 5' to 3' exonuclease activity are useful for the same procedures 25 techniques as their respective non-mutated polymerases except for procedures and techniques which require 5' to 3' exonuclease activity such as the homogeneous assay technique discussed below. Moreover, the mutated the polymerases of present invention 30 oftentimes result in more efficient performance of the procedures and techniques due to the reduction or elimination of the inherent 5' to 3' exonuclease activity.

 and <u>Taf</u> DNA polymerases. In the table below, and throughout the specification, deletion mutations are inclusive of the numbered nucleotides or amino acids which define the deletion.

5	DNA <u>Polymerase</u>	Mutation	Mutant <u>Designation</u>
10	<u>Taq</u>	G(137) to A in nucleotide SED ID NO:1	pRDA3-2
		Gly (46) to Asp in amino acid SEQ ID NO:2	ASP46 Tag
15		Deletion of nucleotides 4-228 of nucleotide SEQ ID NO:1	pTAQd2-76
20		Deletion of amino acids 2-76 of amino acid SEQ ID NO:2	MET-ALA 77 Tag
25		Delection of nucleotides 4-138 of nucleotide SEQ ID NO:1	pTAQd2-46
		Deletion of amino acids 2-46 of amino acid SEQ ID NO:2	MET-PHE 47 Tag
30		Deletion of nucleotides 4-462 of nucleotide SEQ ID NO:1	pTAQd2-155
35		Deletion of amino acids 2-154 of amino acid SEQ ID NO:2	MET-VAL 155 Tag
40		Deletion of nucleotides 4-606 of nucleotide SEQ ID NO:1	pTAQd2-202
45		Deletion of amino acids 2-202 of amino acid SEQ ID NO:2	MET-THR 203 Tag
		Deletion of nucleotides 4-867 of nucleotide SEQ ID NO:1	pLSG8
50			

5	Deletion of amino acids 2-289 of amino acid SEQ ID NO:2	MET-SER 290 Tag (Stoffel fragment)
Tmá	G(110) to A in nucleotide SEQ ID NO:3	•
10	Gly (37) to Asp in amino acid SEQ ID NO:4	ASP37 <u>Tma</u>
15	Deletion of nucleotides 4-131 of nucleotide SEQ ID NO:3	pTMAd2-37
13	Deletion of amino acids 2-37 of amino acid SEQ ID NO:4	MET-VAL 38 Tma
20	Deletion of nucleotides 4-60 of nucleotide SEQ ID NO:3	pTMAd2-20
25	Deletion of amino acids 2-20 of amino acid SEQ ID NO:4	MET-ASP 21 <u>Tma</u>
30	Deletion of nucleotides 4-219 of nucleotide SEQ ID No:3	pTMAd2-73
35	Deletion of amino acids 2-73 amino acid SEQ ID NO: 4	MET-GLU 74 Tma
33	Deletion of nucleotides 1-417 of nucleotide SEQ ID No:3	pTMA16
40	Deletion of amino acids 1-139 of amino acid SEQ ID NO:4	MET 140 Tma
45	Deletion of nucleotides 1-849 of nucleotide SEQ ID No:3	pTMA15
50	Deletion of amino acids 1-283 of amino acid SEQ ID No:4	MET 284 Tma
Tsps17	G(128) to A in nucleotide SEQ ID NO:5	

	Gly (43) to Asp in amino acid SEQ ID NO:6	ASP43 Tsps17
5	Deletion of nucleotides 4-129 of nucleotide SEQ ID NO:5	psPsd2-43
10	Deletion of amino acids 2-43 of amino acid SEQ ID NO:6	MET-PHE 44 Tsps17
	Deletion of nucleotides 4-219 of nucleotide SEQ ID NO:5	pSPSd2-73
15	Deletion of amino acids 2-73 of amino acid SEQ ID NO:6	MET-ALA 74 Tsps17
20	Deletion of nucleotides 4-453 of nucleotide SEQ ID NO:5	pSPSd2-151
25	Deletion of amino acids 2-151 of amino acid SEQ ID NO:6	MET-LEU 152 Tsps17
30	Deletion of nucleotides 4-597 of nucleotide SEQ ID NO:5	pSPSd2-199
	Deletion of amino acids 2-199 of amino acid SEQ ID NO:6	MET-THR 200 Tsps17
35	Deletion of nucleotides 4-861 of nucleotide SEQ ID NO:5	pSPSA288
40	Deletion of amino acids 2-287 of amino acid SEQ ID NO:6	MET-ALA 288 Tsps 17
<u>TZ05</u>	G(137) to A in nucleotide SEQ ID NO:7	
	Gly (46) to Asp in amino acid SEQ ID NO:8	ASP46 <u>TZ05</u>
50	Deletion of nucleotides 4-138 of nucleotide SEQ ID NO:7	pZ05d2-46

•	Deletion of amino acids 2-46 of amino acid SEQ ID NO:8	MET-PHE 47 TZ05
5	Deletion of nucleotides 4-231 of nucleotide SEQ ID NO:7	pZ05d2-77
10	Deletion of amino acids 2-77 of amino acid SEQ ID NO:8	MET-ALA 78 TZ05
15	Deletion of nucleotides 4-475 of nucleotide SEQ ID NO:7	pZ05d2-155
20	Deletion of amino acids 2-155 of amino acid SEQ ID NO:8	MET-VAL 156 TZ05
	Deletion of nucleotides 4-609 of nucleotide SEQ ID NO:7	pZ05d2-203
25	Deletion of amino acids 2-203 of amino acid SEQ ID NO:8	MET-THR 204 <u>TZ05</u>
30	Deletion of nucleotides 4-873 of nucleotide SEQ ID NO:7	p <b>Z05A29</b> 2
35	Deletion of amino acids 2-291 of amino acid SEQ ID NO:8	MET-ALA 292 <u>TZ05</u>
<u>Tth</u>	G(137) to A in nucleotide SEQ ID NO:9	
40	Gly (46) to Asp in amino acid SEQ ID NO:10	ASP46 Tth
45	Deletion of nucleotides 4-138 of nucleotide SEQ ID NO:9	pTTHd2-46
50	Deletion of amino acids 2-46 of amino acid SEQ ID NO:10	MET-PHE 47 Tth
50	Deletion of nucleotides 4-231 of nucleotide	pTTHd2-77

SEQ ID NO:9

	Deletion of amino acids 2-77 of amino acid SEQ ID NO:10	MET-ALA 78 <u>Tth</u>
5	Deletion of nucleotides 4-465 of nucleotide SEQ ID NO:9	pTTHd2-155
10	Deletion of amino acids 2-155 of amino acid SEQ ID NO:10	MET-VAL 156 <u>Tth</u>
15	Deletion of nucleotides 4-609 of nucleotide SEQ ID NO:9	pTTHd2-203
	Deletion of amino acids 2-203 of amino acid SEQ ID NO:10	MET-THR 204 Tth
20	Deletion of nucleotides 4-873 of nucleotide SEQ ID NO:9	рТТНА292
25	Deletion of amino acids 2-291 of amino acid SEQ ID NO:10	MET-ALA 292 Tth
<u>Taf</u> 30	G(110) to A and A(111) to T in nucleotide SEQ ID NO:11	
	Gly (37) to Asp in amino acid SEQ ID NO:12	ASP37 <u>Taf</u>
35	Deletion of nucleotides 4-111 of nucleotide SEQ ID NO:11	pTAFd2-37
40	Deletion of amino acids 2-37 of amino acid SEQ ID NO:12	MET-LEU 38 Taf
45	Deletion of nucleotides 4-279 of nucleotide SEQ ID NO:11	pTAF09
50	Deletion of amino acids 2-93 amino acid SEQ ID NO:12	MET-TYR 94 <u>Taf</u>

25

	Deletion of nucleotides 4-417 of nucleotide SEQ ID NO:11	pTAF11
5	Deletion of amino acids 2-139 of amino acid SEQ ID NO:12	MET-GLU 140 Taf
10	Deletion of nucleotides 4-609 of nucleotide SEQ ID NO:11	pTAFd2-203
15	Deletion of amino acids 2-203 of amino acid SEQ ID NO:12	MET-THR 204 Taf
20	Deletion of nucleotides 4-852 of nucleotide SEQ ID NO:11	pTAFI285
20	Deletion of amino acids 2-284 of amino acid SEQ ID NO:12	MET-ILE 285 Taf

Thermostable DNA Polymerases With Enhanced 5' to 3' Exonuclease Activity

Another aspect of the present invention involves 30 the generation of thermostable DNA polymerases which exhibit enhanced or increased 5' to 3' exonuclease activity that over of their respective polymerases. The thermostable DNA polymerases of the present invention which have increased or enhanced 5' 35 to 3' exonuclease activity are particularly useful in system the homogeneous assay described in PCT application No. 91/05571 filed August 6, 1991, which is incorporated herein by reference. Briefly, this system is a process for the detection of a target amino acid 40 sequence in a sample comprising:

(a) contacting a sample comprising single-stranded nucleic acids with an oligonucleotide containing a sequence complementary to a region of the target 45 nucleic acid and a labeled oligonucleotide containing a

sequence complementary to a second region of the same target nucleic acid strand, but not including the acid sequence defined by the first nucleic oligonucleotide, to create a mixture of duplexes during 5 hybridization conditions, wherein the duplexes comprise to the nucleic acid annealed target oligonucleotide and to the labeled oligonucleotide such that the 3' end of the first oligonucleotide is adjacent to the 5' end of the labeled oligonucleotide;

- 10 (b) maintaining the mixture of step (a) with a template-dependent nucleic acid polymerase having a 5' to 3' nuclease activity under conditions sufficient to permit the 5' to 3' nuclease activity of the polymerase to cleave the annealed, labeled oligonucleotide and 15 release labeled fragments; and
  - (c) detecting and/or measuring the release of labeled fragments.

homogeneous assay system is one which 20 generates signal while the target sequence is minimizing the post-amplification thus, handling of the amplified product which is common to Furthermore, a particularly assay systems. preferred use of the thermostable DNA polymerases with 25 increased 5' to 3*'* exonuclease activity is system homogeneous assay which PCR technology. This particular assay system involves:

- (a) providing to a PCR assay containing said 30 sample, at least one labeled oligonucleotide containing a sequence complementary to a region of the target nucleic acid, wherein said labeled oligonucleotide anneals within the target nucleic acid sequence bounded by the oligonucleotide primers of step (b);
- 35 (b) providing a set of oligonucleotide primers, wherein a first primer contains a sequence

complementary to a region in one strand of the target nucleic acid sequence and primes the synthesis of a complementary DNA strand, and a second primer contains a sequence complementary to a region in a second strand of the target nucleic acid sequence and primes the synthesis of a complementary DNA strand; and wherein each oligonucleotide primer is selected to anneal to its complementary template upstream of any labeled oligonucleotide annealed to the same nucleic acid strand;

- (c) amplifying the target nucleic acid sequence employing a nucleic acid polymerase having 5' to 3' nuclease activity as a template-dependent polymerizing agent under conditions which are permissive for PCR 15 cycling steps of (i) annealing of primers and labeled oligonucleotide to a template nucleic acid sequence contained within the target region, and (ii) extending the primer, wherein said nucleic acid polymerase synthesizes a primer extension product while the 5' to 20 3' nuclease activity of the nucleic acid polymerase simultaneously releases labeled fragments from the annealed duplexes comprising labeled oligonucleotide and its complementary template nucleic acid sequences, thereby creating detectable labeled fragments; and
- 25 (d) detecting and/or measuring the release of labeled fragments to determine the presence or absence of target sequence in the sample.

The increased 5' to 3' exonuclease activity of the 30 thermostable DNA polymerases of the present invention when used in the homogeneous assay systems causes the cleavage of mononucleotides or small oligonucleotides from an oligonucleotide annealed to its larger, complementary polynucleotide. In order for cleavage to 35 occur efficiently, an upstream oligonucleotide must also be annealed to the same larger polynucleotide.

this end of upstream oligonucleotide provides the initial binding site for the nucleic acid polymerase. As soon as the bound polymerase encounters the 5' end of the downstream oligonucleotide, the cleave mononucleotides 5 polymerase can or small oligonucleotides therefrom.

The two oligonucleotides can be designed such that they anneal in close proximity on the complementary target nucleic acid such that binding of the nucleic 10 acid polymerase to the 3' end of the upstream oligonucleotide automatically puts it in contact with the 5' end of the downstream oligonucleotide. This process, because polymerization is not required to bring the nucleic acid polymerase into position to 15 accomplish the cleavage, is called "polymerization-independent cleavage".

Alternatively, if the two oligonucleotides anneal to more distantly spaced regions of the template nucleic acid target, polymerization must occur before 20 the nucleic acid polymerase encounters the 5' end of the downstream oligonucleotide. As the polymerization continues, the polymerase progressively cleaves mononucleotides or small oligonucleotides from the 5' end of the downstream oligonucleotide. This cleaving 25 continues until the remainder of the downstream oligonucleotide has been destabilized to the extent that it dissociates from the template molecule. process is called "polymerization-dependent cleavage".

attachment of label to the downstream 30 oligonucleotide permits the detection of the cleaved mononucleotides and small oligonucleotides. Subsequently, any of several strategies may be employed to distinguish the uncleaved labelled oligonucleotide from the cleaved fragments thereof. In this manner, 35 nucleic acid samples which contain sequences complementary to the upstream and downstream

oligonucleotides can be identified. Stated labelled differently, а oligonucleotide concomittantly with the primer at the start of PCR, and the signal generated from hydrolysis of the labelled 5 nucleotide(s) of the probe provides a means detection of the target sequence during its amplification.

In the homogeneous assay system process, a sample is provided which is suspected of containing 10 particular oligonucleotide sequence of interest, the "target nucleic acid". The target nucleic contained in the sample may first be transcribed into CDNA, if necessary, and then denatured, using any suitable denaturing 15 including physical, chemical, or enzymatic means, which are known to those of skill in the art. A preferred physical means for strand separation involves heating nucleic acid until it is completely (>99%) denatured. Typical heat denaturation involves 20 temperatures ranging from about 80°C to about 105°C; for times ranging from a few seconds to minutes. As an alternative to denaturation, the target nucleic acid may exist in a single-stranded form in the sample, such as, for example, single-stranded RNA or DNA viruses.

denatured nucleic acid strands 25 The incubated with preselected oligonucleotide primers and labeled oligonucleotide (also referred to herein as "probe") under hybridization conditions, conditions which enable the binding of the primers and probes to 30 the single nucleic acid strands. As known in the art, the primers are selected so that their relative positions along a duplex sequence are such that an extension product synthesized from one primer, when the extension product is separated from its template

(complement), serves as a template for the extension of the other primer to yield a replicate chain of defined length.

Because the complementary strands are longer than 5 either the probe or primer, the strands have more points of contact and thus a greater chance of finding each other over any given period of time. A high molar excess of probe, plus the primer, helps tip the balance toward primer and probe annealing rather than template 10 reannealing.

The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. exact length The composition of the primer will depend on many factors, 15 including temperature of the annealing reaction, source and composition of the primer, proximity of the probe annealing site to the primer annealing site, and ratio For example, depending of primer:probe concentration. target sequence, complexity of the the 20 oligonucleotide primer typically contains about 15-30 nucleotides, although a primer may contain more or The primers must be sufficiently fewer nucleotides. complementary to anneal to their respective strands selectively and form stable duplexes.

The primers used herein are selected to be "substantially" complementary to the different strands of each specific sequence to be amplified. The primers need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize selectively to their respective strands.

Non-complementary bases or longer sequences can be interspersed into the primer or located at the ends of the primer, provided the primer retains sufficient complementarity with a template strand to form a stable

duplex therewith. The non-complementary nucleotide sequences of the primers may include restriction enzyme sites.

In the practice of the homogeneous assay system, 5 the labeled oligonucleotide probe must be first annealed to a complementary nucleic acid before the nucleic acid polymerase encounters this duplex region, thereby permitting the 5' to 3' exonuclease activity to cleave and release labeled oligonucleotide fragments.

10 enhance the likelihood that the oligonucleotide will have annealed to a complementary nucleic acid before primer extension polymerization reaches this duplex region, or before the polymerase attaches to upstream oligonucleotide the in 15 polymerization-independent process, variety of techniques may be employed. For the polymerizationdependent process, one can position the probe so that the 5'-end of the probe is relatively far from the 3'-end of the primer, thereby giving the probe more 20 time to anneal before primer extension blocks the probe binding site. Short primer molecules generally require lower temperatures to form sufficiently stable hybrid complexes with the target nucleic acid. Therefore, the labeled oligonucleotide can be designed to be longer 25 than the primer so that the labeled oligonucleotide anneals preferentially to the target higher temperatures relative to primer annealing.

One can also use primers and labeled oligonucleotides having differential thermal 30 stability. For example, the nucleotide composition of the labeled oligonucleotide can be chosen to have greater G/C content and, consequently, greater thermal stability than the primer. In similar fashion, one can incorporate modified nucleotides into the probe, which

modified nucleotides contain base analogs that form more stable base pairs than the bases that are typically present in naturally occurring nucleic acids.

Modifications of the probe that may facilitate 5 probe binding prior to primer binding to maximize the efficiency of the present assav include positively incorporation of charged or phosphodiester linkages in the probe to decrease the repulsion of the polyanionic backbones of the probe and 10 target (see Letsinger et al., 1988, J. Amer. Chem. Soc. 110:4470); incorporation of the alkylated halogenated bases, such as 5-bromouridine, in the probe increase base stacking; the incorporation ribonucleotides into the probe to force 15 probe:target duplex into an "A" structure, which has increased base stacking; and the substitution 2,6-diaminopurine (amino adenosine) for some or all of adenosines in the probe. In preparing such modified probes of the invention, one should recognize 20 that the rate limiting step of duplex formation is "nucleation", the formation of a single base pair, and therefore, altering the biophysical characteristic of a portion of the probe, for instance, only the 3' or 5' terminal portion, can suffice to achieve the desired In addition, because the 3' terminal portion of the probe (the 3' terminal 8 to 12 nucleotides) dissociates following exonuclease degradation of the 5' terminus by the polymerase, modifications of the 3' terminus can be made without concern about interference 30 with polymerase/nuclease activity.

The thermocycling parameters can also be varied to take advantage of the differential thermal stability of the labeled oligonucleotide and primer. For example, following the denaturation step in thermocycling, an 35 intermediate temperature may be introduced which is permissible for labeled oligonucleotide binding but not

primer binding, and then the temperature is further reduced to permit primer annealing and extension. should note, however, that probe cleavage need only occur in later cycles of the PCR process for suitable Thus, one could set up the reaction mixture so that even though primers initially preferentially to probes, primer concentration reduced through primer extension so that, in later cycles, probes bind preferentially to primers.

To favor binding of the labeled oligonucleotide 10 before the primer, a high molar excess of labeled oligonucleotide to primer concentration can also be this embodiment, labeled oligonucleotide concentrations are typically in the range of about 2 to 15 20 times higher than the respective concentration, which is generally  $0.5 - 5 \times 10^{-7} M$ . of skill recognize that oligonucleotide concentration, length, and base composition are each important factors that affect the  $\boldsymbol{T}_{\boldsymbol{m}}$  of any particular 20 oligonucleotide in a reaction mixture. Each of these factors can be manipulated to create a thermodynamic bias to favor probe annealing over primer annealing.

Of course, the homogeneous assay system can be applied to systems that do not involve amplification. 25 In fact, the present invention does not even require that polymerization occur. One: advantage of polymerization-independent process lies the elimination of the need for amplification of the target In the absence of primer extension, sequence. 30 target nucleic acid is substantially single-stranded. Provided the primer and labeled oligonucleotide are adjacently bound to the target nucleic acid, sequential rounds of oligonucleotide annealing and cleavage of labeled fragments can occur. Thus, a sufficient amount 35 of labeled fragments can be generated, making detection

possible in the absence of polymerization. As would be

appreciated by those skilled in the art, the signal generated during PCR amplification could be augmented by this polymerization-independent activity.

addition to the homogeneous assay systems 5 described above, the thermostable DNA polymerases of with enhanced to invention the present useful other activity are also exonuclease transcription systems, such as the amplification amplification system, in which one of the PCR primers 10 encodes a promoter that is used to make RNA copies of In similar fashion, the present the target sequence. invention can be used in a self-sustained sequence replication (3SR) system, in which a variety of enzymes are used to make RNA transcripts that are then used to 15 make DNA copies, all at a single temperature. incorporating a polymerase with 5' to 3' exonuclease activity into a ligase chain reaction (LCR) system, together with appropriate oligonucleotides, one can employ the present invention to detect 20 products.

Also, just as 5' to 3' exonuclease deficient thermostable DNA polymerases are useful in PLCR, other thermostable DNA polymerases which have 5' to 3' exonuclease activity are also useful in PLCR under 25 different circumstances. Such is the case when the 5' primer in PLCR tail of the downstream DNA. Such to the target non-complementary non-complementarity causes a forked structure where the 5' end of the upstream primer would normally anneal to 30 the target DNA.

Thermostable ligases cannot act on such forked structures. However, the presence of 5' to 3' exonuclease activity in the thermostable DNA polymerase will cause the excision of the forked 5' tail of the 35 upstream primer, thus permitting the ligase to act.

WO 92/06200

-74-

PCT/US91/07035

and techniques which The same processes described above as effective for preparing thermostable DNA polymerases with attenuated 5' to 3' exonuclease activity are also effective for preparing 5 thermostable DNA polymerases with enhanced 5' to 3' exonuclease activity. As described above, processes include such techniques as site-directed mutagenesis, deletion mutagenesis and "domain shuffling".

Of particular usefulness preparing 10 in thermostable DNA polymerases with enhanced 5' to 3' "domain shuffling" exonuclease activity is the technique described above. To briefly summarize, this technique involves the cleavage of a specific domain of 15 a polymerase which is recognized as coding for a very active to 3′ exonuclease activity polymerase, and then transferring that domain into the of a second thermostable appropriate area polymerase gene which encodes a lower level or no 5' to The desired domain may 20 3' exonuclease activity. replace a domain which encodes an undesired property of the second thermostable DNA polymerase or be added to the nucleotide sequence of the second thermostable DNA polymerase.

A particular "domain shuffling" example is set 25 forth above in which the Tma DNA polymerase coding sequence comprising codons about 291 through 484 is substituted for the Tag DNA polymerase I codons 289 This substitution yields 422. through 30 thermostable DNA polymerase containing the 5' to 3' Tag DNA polymerase (codons exonuclease domain of 1-289), the 3' to 5' exonuclease domain of Tma DNA polymerase (codons 291-484) and the DNA polymerase polymerase domain of Tag DNA (codons 35 However, those skilled in the art will recognize that other substitutions can be made in order to construct a

thermostable DNA polymerase with certain desired characteristics such as enhanced 5' to 3' exonuclease activity.

The following examples are offered by way of 5 illustration only and are by no means intended to limit the scope of the claimed invention. In these examples, all percentages are by weight if for solids and by volume if for liquids, unless otherwise noted, and all temperatures are given in degrees Celsius.

10

15

### Example 1

Preparation of a 5' to 3' Exonuclease Mutant of <u>Tag</u> DNA Polymerase by Random Mutagenesis PCR of the Known 5' to 3' Exonuclease Domain

### Preparation of Insert

Plasmid pLSG12 was used as a template for PCR.

20 This plasmid is a <u>HindIII</u> minus version of pLSG5 in which the <u>Tag</u> polymerase gene nucleotides 616 - 621 of SEQ ID NO:1 were changed from AAGCTT to AAGCTG. This change eliminated the <u>HindIII</u> recognition sequence within the <u>Tag</u> polymerase gene without altering encoded protein sequence.

Using oligonucleotides MK61 (AGGACTACAACTGCCACACACC) (SEQ ID NO:21) and RA01 (CGAGGCGCCCAGGCCCCAGGAGATCTACC-AGCTCCTTG) (SEQ ID NO:22) as primers and pLSG12 as the template, PCR was conducted to amplify a 384 bp 30 fragment containing the ATG start of the Tag polymerase gene, as well as an additional 331 bp of coding sequence downstream of the ATG start codon.

A 100  $\mu l$  PCR was conducted for 25 cycles utilizing the following amounts of the following agents and 35 reactants:

50 pmol of primer MK61 (SEQ ID NO:21);
50 pmol of primer RA01 (SEQ ID NO:22);
50 μM of each dNTP;
10 mM Tris-HCl, pH 8.3;
5 50 mM KCl;
1.5 mM MgCl<sub>2</sub>;
75.6 pg pLSG12;
2.5 units AmpliTaq DNA polymerase.

- The PCR reaction mixture described was placed in a Perkin-Elmer Cetus Thermocycler and run through the following profile. The reaction mixture was first ramped up to 98°C over 1 minute and 45 seconds, and held at 98°C for 25 seconds. The reaction mixture was 15 then ramped down to 55°C over 45 seconds and held at that temperature for 20 seconds. Finally, the mixture was ramped up to 72°C over 45 seconds, and held at 72°C for 30 seconds. A final 5 minute extension occurred at 72°C.
- The PCR product was then extracted with chloroform and precipitated with isopropanol using techniques which are well known in the art.

A 300 ng sample of the PCR product was digested with 20 U of <u>Hin</u>dIII (in 30 µl reaction) for 2 hours at 25 37°C. Then, an additional digestion was made with 8 U of <u>Bss</u>HII for an 2 hours at 50°C. This series of digestions yielded a 330 bp fragment for cloning.

A vector was prepared by digesting 5.3  $\mu$ g of pLSG12 with 20 U <u>Hin</u>dIII (in 40  $\mu$ l) for 2 hours at 37°C. This 30 digestion was followed by addition of 12 U of <u>Bss</u>HII and incubation for 2 hours at 50°C.

The vector was dephosphorylated by treatment with CIAP (calf intestinal alkaline phosphatase), specifically 0.04 U CIAP for 30 minutes at 30°C. Then,

4  $\mu l$  of 500 mM EGTA was added to the vector preparation to stop the reaction, and the phosphatase was inactivated by incubation at 65°C for 45 minutes.

225 ng of the phosphatased vector described above 5 was ligated at a 1:1 molar ratio with 10 ng of the PCR-derived insert.

Then, DG116 cells were transformed with one fifth of the ligation mixture, and ampicillin-resistant transformants were selected at 30°C.

10 Appropriate colonies were grown overnight at 30°C to OD<sub>600</sub> 0.7. Cells containing the P<sub>L</sub> vectors were induced at 37°C in a shaking water bath for 4, 9, or 20 hours, and the preparations were sonicated and heat treated at 75°C in the presence of 0.2 M ammonium 15 sulfate. Finally, the extracts were assayed for polymerase activity and 5' to 3' exonuclease activity.

The 5' to 3' exonuclease activity was quantified utilizing the 5' to 3' exonuclease assay described Specifically, the synthetic 3' phosphorylated 20 oligonucleotide probe (phosphorylated to preclude (GATCGCTGCGCGTAACCACCA-BW33 polymerase extension) pmol) NO:13) (100 (SEQ ID CACCCGCCGCGCp)  $^{32}$ P-labeled at the 5' end with gamma-[ $^{32}$ P] ATP (3000 Ci/mmol) and T4 polynucleotide kinase. The reaction 25 mixture was extracted with phenol:chloroform:isoamyl alcohol, followed by ethanol precipitation. 32P-labeled oligonucleotide probe was redissolved in 100 µl of TE buffer, and unincorporated ATP was removed by gel filtration chromatography on a Sephadex G-50 30 spin column. Five pmol of 32P-labeled BW33 probe, was annealed to 5 pmol of single-strand M13mp10w DNA, in the presence of 5 pmol of the synthetic oligonucleotide primer BW37 (GCGCTAGGGCGCTGGCAAGTGTAGCGGTCA) NO:14) in a 100  $\mu$ l reaction containing 10 mM Tris-HCl 35 (pH 8.3), 50 mM KCl, and 3 mM MgCl<sub>2</sub>. The annealing mixture was heated to 95°C for 5 minutes, cooled to

70°C over 10 minutes, incubated at 70°C for an additional 10 minutes, and then cooled to 25°C over a 30 minute period in a Perkin-Elmer Cetus DNA thermal cycler. Exonuclease reactions containing 10  $\mu$ l of the 5 annealing mixture were pre-incubated at 70°C for 1 The thermostable DNA polymerase preparations minute. the invention (approximately 0.3 U of enzyme activity) were added in a 2.5 µl volume to the pre-incubation reaction, and the reaction mixture was 10 incubated at 70°C. Aliquots (5 μl) were removed after 1 minute and 5 minutes, and stopped by the addition of 1 µl of 60 mM EDTA. The reaction products were analyzed by homochromatography and exonuclease activity quantified following autoradiography. 15 Chromatography was carried out in a homochromatography mix containing 2% partially hydrolyzed yeast RNA in 7M urea on Polygram CEL 300 DEAE cellulose thin layer The presence of 5' to 3' chromatography plates. exonuclease activity resulted in the generation of 20 small <sup>32</sup>P-labeled oligomers, which migrated up the TLC and were easily differentiated autoradiogram from undegraded probe, which remained at the origin.

The clone 3-2 had an expected level of polymerase 25 activity but barely detectable 5' to 3' exonuclease activity. This represented a greater than 1000-fold reduction in 5' to 3' exonuclease activity from that present in native <u>Tag</u> DNA polymerase.

This clone was then sequenced and it was found that 30 G (137) was mutated to an A in the DNA sequence. This mutation results in a Gly (46) to Asp mutation in the amino acid sequence of the <u>Tag</u> DNA polymerase, thus yielding a thermostable DNA polymerase of the present invention with significantly attenuated 5' to 3' 35 exonuclease activity.

WO 92/06200 PCT/US91/07035

-79-

The recovered protein was purified according to the <u>Tag</u> DNA polymerase protocol which is taught in Serial No. 523,394 filed May 15, 1990, incorporated herein by reference.

5

### Example 2

# Construction of Met 289 (∆289) 544 Amino Acid Form of Tag Polymerase

10

As indicated in Example 9 of U.S. Serial No. 523,394, filed May 15, 1990, during a purification of native Tag polymerase an altered form of Tag polymerase was obtained that catalyzed the template dependent 15 incorporation of dNTP at 70°C. This altered form of polymerase was immunologically related to approximate 90 kd form of purified native Tag polymerase but was of lower molecular weight. mobility, relative to BSA and ovalbumin following 20 SDS-PAGE electrophoresis, the apparent molecular weight of this form is approximately 61 kd. This altered form of the enzyme is not present in carefully prepared crude extracts of Thermus aquaticus cells as determined by SDS-PAGE Western blot analysis or <u>in situ</u> DNA 25 polymerase activity determination (Spanos, Α., and Hubscher, U. (1983) Meth. Enz. 91:263-277) following SDS-PAGE gel electrophoresis. This form appears to be a proteolytic artifact that may arise during sample handling. This lower molecular weight 30 purified to homogeneity and subjected to N-terminal sequence determination on an ABI automated gas phase Comparison of the obtained N-terminal sequencer. sequence with the predicted amino acid sequence of the Tag polymerase gene (SEQ ID NO:1) indicates this 35 shorter form arose as a result of proteolytic cleavage between Glu(289) and Ser(290).

To obtain a further truncated form of a polymerase gene that would direct the synthesis of a 544 amino acid primary translation production plasmids pSYC1578 and the complementary synthetic 5 oligonucleotides DG29 (5'-AGCTTATGTCTCCAAAAGCT) (SEQ ID NO:23) and DG30 (5'-AGCTTTTGGAGACATA) (SEQ ID NO:24) were used. Plasmid pFC54.t was digested to completion with <u>HindIII</u> and <u>Bam</u>HI. Plasmid pSYC1578 was digested with <a href="mailto:BStXI">BStXI</a> (at nucleotides 872 to 883 of SEQ ID NO:1) 10 and treated with  $\underline{E}$ .  $\underline{coli}$  DNA polymerase I Klenow fragment in the presence of all 4 dNTPs to remove the 4 nucleotide 3′ cohesive end and generate CTG-terminated duplex blunt end encoding Leu294 in the Tag polymerase sequence (see Tag polymerase SEQ ID NO:1 15 nucleotides 880-882). The DNA sample was digested to completion with  $\underline{\mathtt{Bql}}\mathtt{II}$  and the approximate 1.6 kb  $\underline{\mathtt{Bst}}\mathtt{XI}$ (repaired)/BglII Tag DNA fragment was purified by agarose gel electrophoresis and electroelution. pFC54.t plasmid digest (0.1 pmole) was ligated with the 20 Tag polymerase gene fragment (0.3 pmole) and annealed nonphosphorylated DG29/DG30 duplex adaptor (0.5 pmole) under sticky ligase conditions at 30 µg/ml, The DNA was diluted to approximately 10 overnight. microgram per ml and ligation continued under blunt end 25 conditions. The ligated DNA sample was digested with <u>Xba</u>I to linearize (inactivate) any IL-2 mutein-encoding ligation products. 80 nanograms of the ligated and digested' DNA was used to transform E. coli K12 strain DG116 to ampicillin resistance. Amp<sup>R</sup> candidates were 30 screened for the presence of an approximate 7.17 kb plasmid which yielded the expected digestion products with EcoRI (4,781 bp + 2,386 bp), PstI (4,138 bp + 3,029 bp),  $\underline{Apa}I$  (7,167 bp) and  $\underline{Hin}dIII/\underline{Pst}I$  (3,400 bp + bp + 738 bp).3,029 E. coli colonies harboring 35 candidate plasmids were screened by single colony immunoblot for the temperature-inducible synthesis of

an approximate 61 kd Tag polymerase related polypeptide. In addition, candidate plasmids were subjected to DNA sequence determination at the 5'  $\lambda P_{\rm L}$ promoter: Taq DNA junction and the 3' Taq DNA: BT cry PRE One of the plasmids encoding the intended sequence and directing the synthesis temperature-inducible 61 kd Tag polymerase related polypeptide was designated pLSG68.

Expression of 61 kDa Tag Pol I. Cultures 10 containing pLSG8 were grown as taught in Serial No. 523,364 and described in Example 3 below. The 61 kDa Tag Pol appears not to be degraded heat-induction at 41°C. After 21 hours at 41°C, a heat-treated crude extract from a culture harboring 15 pLSG8 had 12,310 units of heat-stable DNA polymerase activity per mg crude extract protein, a 24-fold increase over an uninduced culture. A heat-treated extract from a 21 hour 37°C-induced pLSG8 culture had 9,503 units of activity per mg crude extract protein. 20 A nine-fold increase in accumulated levels of Tag Pol I was observed between a 5 hour and 21 hour induction at

- was observed between a 5 hour and 21 hour induction at 37°C and a nearly four-fold increase between a 5 hour and 21 hour induction at 41°C. The same total protein and heat-treated extracts were analyzed by SDS-PAGE.
- 25 20 μg crude extract protein or heat-treated crude extract from 20 μg crude extract protein were applied to each lane of the gel. The major bands readily apparent in both the 17°C and 41°C, 21 hour-induced total protein lanes are equally intense as their
- 30 heat-treated counterparts. Heat-treated crude extracts from 20 μg of total protein from 37°C and 41°C, 21 hour samples contain 186 units and 243 units of thermostable DNA polymerase activity, respectively. To determine the usefulness of 61 kDa <u>Tag</u> DNA polymerase in PCR, PCR
- 35 assays were performed using heat-treated crude extracts from induced cultures of pLSG8. Heat-treated crude

extract from induced cultures of pLSG5 were used as the source of full-length <u>Tag</u> Pol I in PCR. PCR product was observed in reactions utilizing 4 units and 2 units of truncated enzyme. There was more product in those 5 PCRs than in anyof the full-length enzyme reactions. In addition, no non-specific higher molecular weight products were visible.

Purification of 61 kDa Tag Pol I. Purification of 61 kDa Tag Pol I from induced pLSG8/DG116 cells 10 proceeded as the purification of full-length Tag Pol I as in Example 12 of U.S. Serial No. 523,394, filed May 15, 1990 with some modifications.

Induced pLSG8/DG116 cells (15.6 g) were homogenized and lysed as described in U.S. Serial No. 523,394, 15 filed May 15, 1990 and in Example 3 below. Fraction I contained 1.87 g protein and 1.047 x 10<sup>6</sup> units of activity. Fraction II, obtained as a 0.2 M ammonium sulfate supernant contained 1.84 g protein and 1.28 x 10<sup>6</sup> units of activity in 74 ml.

20 Following heat treatment, Polymin P (pH 7.5) was added slowly to 0.7%. Following centrifugation, the supernant, Fraction III contained 155 mg protein and 1.48 x 10<sup>6</sup> units of activity.

Fraction III was loaded onto a 1.15 x 3.1 cm (3.2 25 ml) phenyl sepharose column at 10 ml/cm²/hour. All of the applied activity was retained on the column. The column was washed with 15 ml of the equilibration buffer and then 5 ml (1.5 column volumes) of 0.1M KCl in TE. The polymerase activity was eluted with 2 M 30 urea in TE containing 20% ethylene glycol. Fractions (0.5 ml each) with polymerase activity were pooled (8.5 ml) and dialyzed into heparin sepharose buffer containing 0.1 M KCl. The dialyzed material, Fraction IV (12.5 ml), contained 5.63 mg of protein and 1.29 x 35 106 units of activity.

Fraction IV was loaded onto a 1.0 ml bed volume heparin sepharose column equilibrated as above. column was washed with 6 ml of the same buffer ( $A_{280}$  to baseline) and eluted with a 15 ml linear 0.1-0.5 M KCl 5 gradient in the same buffer. Fractions (0.15 ml) eluting between 0.16 and 0.27 M KCl were analyzed by A minor (<1%) contaminating approximately 47 kDa protein copurified with 61 kDa <u>Taq</u> Pol Fractions eluting between 0.165 and 0.255 M KCl were 10 pooled (2.5 ml) and diafiltered on a Centricon 30 into 2.5% storage buffer. Fraction contained 2.8 mg of protein and 1.033 x 106 units of 61 kDa Tag Pol I.

PCR Using Purified 61 kDa Tag Pol I. PCR reactions 15 (50 µl) containing 0.5 ng lambda DNA, 10 pmol each of two lambda-specific primers, 200  $\mu M$  each dNTPs, 10 mMTris-Cl, pH 8.3, 3 mM MgCl2, 10 mM KCl and 3.5 units of 61 kDa Tag Pol I were performed. As a comparison, PCR reactions were performed with 1.25 units of full-length 20 Tag Pol I, as above, with the substitution of 2 mM MgCl<sub>2</sub> and 50 mM KCl. Thermocycling conditions were 1 minute at 95°C and 1 minute at 60°C for 23 cycles, with a final 5 minute extension at 75°C. The amount of DNA per reaction was quantitated by the Hoechst fluorescent 25 dye assay. 1.11 μg of product was obtained with 61 kDa Tag Pol I (2.2 x 105-fold amplification), as compared with 0.70  $\mu g$  of DNA with full-length Tag Pol I (1.4  $\times$ 105-fold amplification).

Thermostability of 61 kDa Taq Pol I. Steady state 30 thermal inactivation of recombinant 94 kDa Taq Pol I and 61 kDa Taq Pol I was performed 97.5°C under buffer conditions mimicking PCR. 94 kDa Taq Pol I has an apparent half-life of approximately 9 minute at 97.5°C, whereas the half-life of 61 kDa Taq Pol I was

approximately 21 minutes. The thermal inactivation of 61 kDa <u>Tag</u> Pol I was unaffected by KCl concentration over a range from 0 to 50 mM.

Yet another truncated <u>Tag</u> polymerase gene contained 5 within the ~2.68 kb <u>HindIII-Asp</u>718 fragment of plasmid pFC85 can be expressed using, for example, plasmid pP<sub>L</sub>N<sub>RBS</sub>ATG, by operably linking the amino-terminal <u>HindIII</u> restriction site encoding the <u>Tag pol</u> gene to an ATG initiation codon. The product of this fusion 10 upon expression will yield an ~70,000-72,000 dalton truncated polymerase.

This specific construction can be made by digesting plasmid pFC85 with <u>Hin</u>dIII and treating with Klenow fragment in the presence of dATP and dGTP. 15 resulting fragment is treated further with Sl nuclease remove any single-stranded extensions and the resulting DNA digested with Asp718 and treated with Klenow fragment in the presence of all four dNTPs. recovered fragment can be ligated using T4 DNA ligase 20 to dephosphorylated plasmid  $\mathtt{pP_LN_{RBS}ATG}$ , which had been digested with <a href="Sac">Sac</a>I and treated with Klenow fragment in the presence of dGTP to construct an ATG blunt end. This ligation mixture can then be used to transform E. <u>coli</u> DG116 and transformants the screened 25 production of Tag polymerase. Expression can confirmed by Western immunoblot analysis and activity analysis.

### Example 3

30

Construction, Expression and Purification of a Truncated 5' to 3' Exonuclease

Deficient Tma Polymerase (MET284)

WO 92/06200 PCT/US91/07035

To express a 5' to 3' exonuclease deficient <u>Tma</u> DNA polymerase lacking amino acids 1-283 of native <u>Tma</u> DNA polymerase the following steps were performed.

Plasmid pTma12-1 was digested with **BspHI** 5 (nucleotide position 848) and <u>HindIII</u> (nucleotide position 2629). A 1781 base pair fragment was isolated by agarose gel purification. To separate the agarose a gel slice containing the desired from the DNA. fragment was frozen at -20°C in a Costar spinex filter 10 unit. After thawing at room temperature, the unit was spun in a microfuge. The filtrate containing the DNA was concentrated in a Speed Vac concentrator, and the DNA was precipitated with ethanol.

The isolated fragment was cloned into plasmid 15 pTmal2-1 digested with NcoI and HindIII. Because NcoI digestion leaves the same cohesive end sequence as digestion with BspHI, the 1781 base pair fragment has the same cohesive ends as the full length fragment excised from plasmid pTmal2-1 by digestion with NcoI and HindIII. The ligation of the isolated fragment with the digested plasmid results in a fragment switch and was used to create a plasmid designated pTmal4.

Plasmid pTma15 was similarly constructed by cloning the same isolated fragment into pTma13. As with 25 pTma14, pTma15 drives expression of a polymerase that lacks amino acids 1 through 283 of native Tma DNA polymerase; translation initiates at the methionine codon at position 284 of the native coding sequence.

Both the pTmal4 and pTmal5 expression plasmids

30 expressed at a high level a biologically active
thermostable DNA polymerase devoid of 5' to 3'
exonuclease activity of molecular weight of about 70
kDa; plasmid pTmal5 expressed polymerase at a higher
level than did pTmal4. Based on similarities with E.

35 coli Pol I Klenow fragment, such as conservation of
amino acid sequence motifs in all three domains that

are critical for 3' to 5' exonuclease activity, distance from the amino terminus to the first domain critical for exonuclease activity, and length of the expressed protein, the shortened form (MET284) of Tma DNA polymerase exhibits 3' to 5' exonuclease or proof-reading activity but lacks 5' to 3' exonuclease activity. Initial SDS activity gel assays and solution assays for 3' to 5' exonuclease activity suggest attenuation in the level of proof-reading activity of the polymerase expressed by E. coli host cells harboring plasmid pTma15.

MET284 Tma DNA polymerase was purified from E. coli strain DG116 containing plasmid pTma15. The seed flask for a 10 L fermentation contained tryptone (20 g/l), 15 yeast extract (10 g/l), NaCl (10 g/l), glucose (10 g/l), ampicillin (50 mg/l), and thiamine (10 mg/l). The seed flask was innoculated with a colony from an agar plate (a frozen glycerol culture can be used). seed flask was grown at 30°C to between 0.5 to 2.0 O.D. 20 (A680). The volume of seed culture inoculated into the fermentor is calculated such that the bacterial concentration is 0.5 mg dry weight/liter. The 10 liter growth medium contained 25 mM  $KH_2PO_4$ , 10 mM  $(NH_4)_2SO_4$ , 4 mM sodium citrate, 0.4 mM FeCl3, 0.04 mM ZnCl2, 0.03 25 mM  $CoCl_2$ , 0.03 mM  $CuCl_2$ , and 0.03 mM  $H_3BO_3$ . following sterile components were added: 4 mM MgSO<sub>4</sub>, g/l glucose, 20 mg/l thiamine, and 50 mq/1ampicillin. The pH was adjusted to 6.8 with NaOH and controlled during the fermentation by added NHAOH. 30 Glucose was continually added by coupling to NH4OH addition. Foaming was controlled by the addition of propylene glycol as necessary, as an antifoaming agent. Dissolved oxygen concentration was maintained at 40%.

The fermentor was inoculated as described above, 35 and the culture was grown at 30°C to a cell density of 0.5 to 1.0 X  $10^{10}$  cells/ml (optical density [A<sub>680</sub>] of

15). The growth temperature was shifted to 38°C to induce the synthesis of MET284 <u>Tma</u> DNA polymerase. The temperature shift increases the copy number of the pTma15 plasmid and simultaneously derepresses the lambda P<sub>L</sub> promoter controlling transcription of the modified <u>Tma</u> DNA polymerase gene through inactivation of the temperature-sensitive cI repressor encoded by the defective prophage lysogen in the host.

The cells were grown for 6 hours to an optical 10 density of 37 (A<sub>680</sub>) and harvested by centrifugation. The cell mass (ca. 95 g/l) was resuspended in an equivalent volume of buffer containing 50 mM Tris-Cl, pH 7.6, 20 mM EDTA and 20% (w/v) glycerol. The suspension was slowly dripped into liquid nitrogen to 15 freeze the suspension as "beads" or small pellets. The frozen cells were stored at -70°C.

To 200 g of frozen beads (containing 100 g wet weight cell) were added 100 ml of 1X TE (50 mM Tris-Cl, pH 7.5, 10 mM EDTA) and DTT to 0.3 mM, PMSF to 2.4 mM, 20 leupeptin to 1  $\mu$ g/ml and TLCK (a protease inhibitor) to The sample was thawed on ice and uniformly 0.2 mM. resuspended in a blender at low speed. suspension was lysed in an Aminco french pressure cell at 20,000 psi. To reduce viscosity, the lysed cell 25 sample was sonicated 4 times for 3 min. each at 50% duty cycle and 70% output. The sonicate was adjusted to 550 ml with 1X TE containing 1 mM DTT, 2.4 mM PMSF, 1 μg/ml leupeptin and 0.2 mM TLCK (Fraction I). addition of ammonium sulfate to 0.3 M, the crude lysate 30 was rapidly brought to 75°C in a boiling water bath and transferred to a 75°C water bath for 15 min. to denature and inactivate  $\underline{E}$ .  $\underline{coli}$  host proteins. heat-treated sample was chilled rapidly to 0°C and incubated on ice for 20 min. Precipitated proteins and

cell membranes were removed by centrifugation at 20,000  $\times$  G for 30 min. at 5°C and the supernatant (Fraction II) saved.

The heat-treated supernatant (Fraction II) 5 treated with polyethyleneimine (PEI) to remove most of the DNA and RNA. Polymin P (34.96 ml of 10% [w/v], pH 7.5) was slowly added to 437 ml of Fraction II at 0°C while stirring rapidly. After 30 min. at 0°C, the sample was centrifuged at 20,000 X G for 30 min. 10 supernatant (Fraction III) was applied at 80 ml/hr to a 100 ml phenylsepharose column (3.2 x 12.5 cm) that had been equilibrated in 50 mM Tris-Cl, pH 7.5, 0.3 M ammonium sulfate, 10 mM EDTA, and 1 mM DTT. The column was washed with about 200 ml of the same buffer (A280 15 to baseline) and then with 150 ml of 50 mM Tris-Cl, pH 7.5, 100 mM KCl, 10 mM EDTA and 1 mM DTT. The MET284 Tma DNA polymerase was then eluted from the column with buffer containing 50 mM Tris-Cl, pH 7.5, 2 M urea, 20% (w/v) ethylene glycol, 10 mM EDTA, and 1 mM DTT, and 20 fractions containing DNA polymerase activity were pooled (Fraction IV).

Fraction IV is adjusted conductivity to a equivalent to 50 mM KCl in 50 mM Tris-Cl, pH 7.5, 1 mM EDTA, and 1 mM DTT. The sample was applied (at 9 25 ml/hr) to a 15 ml heparin-sepharose column that had been equilibrated in the same buffer. The column was washed with the same buffer at ca. 14 ml/hr (3.5 column volumes) and eluted with a 150 ml 0.05 to 0.5 M KCl gradient in the same buffer. The DNA polymerase 30 activity eluted between 0.11-0.22 M KCl. Fractions containing the pTma15 encoded modifed Tma DNA polymerase are pooled, concentrated, and diafiltered against 2.5% storage buffer (50 mM Tris-Cl, pH 8.0, 250 mM KCl, 0.25 mM EDTA, 2.5 mM DTT, and 0.5% Tween 20), 35 subsequently mixed with 1.5 volumes of sterile 80% (w/v) glycerol, and stored at -20°C. Optionally, the

heparin sepharose-eluted DNA polymerase or the phenyl sepharose-eluted DNA polymerase can be dialyzed or adjusted to a conductivity equivalent to 50 mM KCl in 50 mM Tris-Cl, pH 7.5, 1 mM DTT, 1 mM EDTA, and 0.2% 5 Tween 20 and applied (1 mg protein/ml resin) to an affigel blue column that has been equilibrated in the same buffer. The column is washed with three to five column volumes of the same buffer and eluted with a 10 column volume KCl gradient (0.05 to 0.8 M) in the same buffer. Fractions containing DNA polymerase activity (eluting between 0.25 and 0.4 M KCl) are pooled, concentrated, diafiltered, and stored as above.

The relative thermoresistance of various DNA polymerases has been compared. At 97.5°C the half-life 15 of native <a href="mailto:Tma">Tma</a> DNA polymerase is more than twice the half-life of either native or recombinant <a href="mailto:Tag">Tag</a> DNA (i.e., AmpliTag) DNA polymerase. Surprisingly, the half-life at 97.5°C of MET284 <a href="mailto:Tma">Tma</a> DNA polymerase is 2.5 to 3 times longer than the half-life of native <a href="mailto:Tma">Tma</a> DNA polymerase.

PCR tubes containing 10 mM Tris-Cl, pH 8.3, and 1.5 mM MgCl<sub>2</sub> (for <u>Tag</u> or native <u>Tma</u> DNA polymerase) or 3 mM MgCl<sub>2</sub> (for MET284 <u>Tma</u> DNA polymerase), 50 mM KCl (for <u>Tag</u>, native <u>Tma</u> and MET284 <u>Tma</u> DNA polymerases) or no 25 KCl (for MET284 <u>Tma</u> DNA polymerase), 0.5 µM each of primers PCR01 and PCR02, 1 ng of lambda template DNA, 200 µM of each dNTP except dCTP, and 4 units of each enzyme were incubated at 97.5°C in a large water bath for times ranging from 0 to 60 min. Samples were 30 withdrawn with time, stored at 0°C, and 5 µl assayed at 75°C for 10 min. in a standard activity assay for residual activity.

Tag DNA polymerase had a half-life of about 10 min. at 97.5°C, while native Tma DNA polymerase had a 35 half-life of about 21 to 22 min. at 97.5°C. Surprisingly, the MET284 form of Tma DNA polymerase had

a significanlty longer half-life (50 to 55 min.) than either <u>Tag</u> or native <u>Tma</u> DNA polymerase. The improved thermoresistance of MET284 <u>Tma</u> DNA polymerase will find applications in PCR, particularly where G+C-rich targets are difficult to amplify because the strand-separation temperature required for complete denaturation of target and PCR product sequences leads to enzyme inactivation.

PCR tubes containing 50 µl of 10 mM Tris-Cl, pH 3 mM MgCl<sub>2</sub>, 200 µM of each dNTP, 0.5 ng bacteriophage lambda DNA, 0.5  $\mu M$  of primer PCR01, 4 units of MET284  $\underline{\text{Tma}}$  DNA polymerase, and 0.5  $\mu\text{M}$  of primer PCR02 or PL10 were cycled for 25 cycles using T<sub>den</sub> of 96°C for 1 min. and T<sub>anneal-extend</sub> of 60°C for 15 2 min. Lambda DNA template, deoxynucleotide stock solutions, and primers PCR01 and PCR02 were part of the PECI GeneAmp kit. Primer PL10 has the sequence: 5'-GGCGTACCTTTGTCTCACGGGCAAC-3' (SEQ ID NO:25) and is complementary to bacteriophage lambda nucleotides 20 8106-8130.

The primers PCR01 and PCR02 amplify a 500 bp product from lambda. The primer pair PCR01 and PL10 amplify 1 а kb product from lambda. amplification with the respective primer sets, 5  $\mu$ l 25 aliquots were subjected to agarose gel electrophoresis and the specific intended product bands visualized with ethidium bromide staining. Abundant levels of product were generated with both primer sets, showing that MET284 Tma DNA polymerase successfully amplified the 30 intended target sequence.

### Example 4

# Expression of Truncated Tma DNA Polymerase

To express a 5' to 3' exonuclease deficient form of <a href="Tma">Tma</a> DNA polymerase which initiates translation at MET 140 the coding region corresponding to amino acids 1 through 139 was deleted from the expression vector. The protocol for constructing such a deletion is 10 similar to the construction described in Examples 2 and 3: a shortened gene fragment is excised and then reinserted into a vector from which a full length fragment has been excised. However, the shortened fragment can be obtained as a PCR amplification product 15 rather than purified from a restriction digest. This methodology allows a new upstream restriction site (or other sequences) to be incorporated where useful.

To delete the region up to the methionine codon at position 140, an <u>Sph</u>I site was introduced into pTma12-1 20 and pTma13 using PCR. A forward primer corresponding to nucleotides 409-436 of <u>Tma</u> DNA polymerase SEQ ID NO:3 (FL63) was designed to introduce an <u>Sph</u>I site just upstream of the methionine codon at position 140. The reverse primer corresponding to the complement of 25 nucleotides 608-634 of <u>Tma</u> DNA polymerase SEQ ID NO:3 (FL69) was chosen to include an <u>Xba</u>I site at position 621. Plasmid pTma12-1 linearized with <u>Sma</u>I was used as the PCR template, yielding an approximate 225 bp PCR product.

Before digestion, the PCR product was treated with 50 μg/ml of Proteinase K in PCR reaction mix plus 0.5% SDS and 5 mM EDTA. After incubating for 30 minutes at 37°C, the Proteinase K was heat inactivated at 68°C for 10 minutes. This procedure eliminated any Tag 35 polymerase bound to the product that could inhibit

subsequent restriction digests. The buffer was changed to a TE buffer, and the excess PCR primers were removed with a Centricon 100 microconcentrator.

The amplified fragment was digested with SphI, then 5 treated with Klenow to create a blunt end at the SphI-cleaved end, and finally digested with XbaI. resulting fragment was ligated with plasmid pTmal3 (pTmal2-1 would have been suitable) that had been digested with NcoI, repaired with Klenow, and then 10 digested with XbaI. The ligation yielded an in-frame coding sequence with the region following the NcoI site (at the first methionine codon of the coding sequence) introduced <u>Sph</u>I site (upstream the methionine codon at position 140) deleted. The 15 resulting expression vector was designated pTmal6.

The primers used in this example are given below and in the Sequence Listing section.

	<u>Primer</u>	SEO ID NO:	Sequence
20			
	FL63	SEQ ID NO:26	5'GATAAAGGCATGCTTCAGCTTGTGAACG
	FL69	SEQ ID NO:27	5'TGTACTTCTCTAGAAGCTGAACAGCAG

# Example 5

25

# Elimination of Undesired RBS in MET140 Expression Vectors

Reduced expression of the MET140 form of Tma DNA 30 polymerase can be achieved by eliminating the ribosome binding site (RBS) upstream of the methionine codon at position 140. The RBS was eliminated be via oligonucleotide site-directed mutagenesis Without changing the amino acid sequence. Taking advantage of 35 the redundancy of the genetic code, one can make changes in the third position of codons to alter the

nucleic acid sequence, thereby eliminating the RBS, without changing the amino acid sequence of the encoded protein.

A mutagenic primer (FL64) containing the modified 5 sequence was synthesized and phosphorylated. Single-stranded pTma09 (a full length clone having an NcoI site) was prepared by coinfecting with the helper phage R408, commercially available from Stratagene. "gapped duplex" of single stranded pTma09 and the large 10 fragment from the PvuII digestion of pBS13+ was created by mixing the two plasmids, heating to boiling for 2 minutes, and cooling to 65°C for 5 minutes. phosphorylated primer was then annealed with "gapped duplex" by mixing, heating to 80°C for 2 15 minutes, and then cooling slowly to room temperature. The remaining gaps were filled by extension with Klenow and the fragments ligated with T4 DNA ligase, both reactions taking place in 200  $\mu M$  of each dNTP and 40  $\mu M$ 

The resulting circular fragment was transformed into DG101 host cells by plate transformations on nitrocellulose filters. Duplicate filters were made and the presence of the correct plasmid was detected by probing with a  $\gamma^{32}$ P-phosphorylated probe (FL65). The vector that resulted was designated pTma19.

ATP in standard salts at 37°C for 30 minutes.

The RBS minus portion from pTma19 was cloned into pTma12-1 via an NcoI/XbaI fragment switch. Plasmid pTma19 was digested with NcoI and XbaI, and the 620 bp fragment was purified by gel electrophoresis, as in 30 Example 3, above. Plasmid pTma12-1 was digested with NcoI, XbaI, and XcmI. The XcmI cleavage inactivates the RBS+ fragment for the subsequent ligation step, which is done under conditions suitable for ligating "sticky" ends (dilute ligase and 40 µM ATP). Finally, 35 the ligation product is transformed into DG116 host cells for expression and designated pTma19-RBS.

The oligonucleotide sequences used in this example are listed below and in the Sequence Listing section.

	<u>Oligo</u>	SEQ ID NO:	Sequence
5			
	FL64	SEQ ID NO:28	5'CTGAAGCATGTCTTTGTCACCGGT-
			TACTATGAATAT
	FL65 .	SEQ ID NO:29	5'TAGTAACCGGTGACAAAG

10 Example 6

## Expression of Truncated Tma DNA Polymerases MET-ASP21 and MET-GLU74

- To effect translation initiation at the aspartic acid codon at position 21 of the <u>Tma</u> DNA polymerase gene coding sequence, a methionine codon is introduced before the codon, and the region from the initial <u>Nco</u>I site to this introduced methionine codon is deleted. Similar to Example 4, the deletion process involved PCR with the same downstream primer described above (FL69) and an upstream primer (FL66) designed to incorporate an <u>Nco</u>I site and a methionine codon to yield a 570 base pair product.
- 25 The amplified product was concentrated with a Centricon-100 microconcentrator to eliminate primers and buffer. The product was concentrated in a Speed Vac concentrator and then resuspended in the digestion mix. The amplified product was digested with 30 Ncol and XbaI. Likewise, pTma12-1, pTma13. pTmal9-RBS was digested with the same two restriction enzymes, and the digested, amplified fragment is ligated with the digested expression vector. The resulting construct has a deletion from the NcoI site upstream of 35 the start codon of the native Tma coding sequence to the

new methionine codon introduced upstream of the aspartic acid codon at position 21 of the native <u>Tma</u> coding sequence.

Similarly, a deletion mutant was created such that 5 translation initiation begins at Glu74, the glutamic acid codon at position 74 of the native <u>Tma</u> coding sequence. An upstream primer (FL67) is designed to introduce a methionine codon and an <u>NcoI</u> site before Glu74. The downstream primer and cloning protocol used 10 are as described above for the MET-ASP21 construct.

The upstream primer sequences used in this example are listed below and in the Sequence Listing section.

	O <u>ligo</u>	SEO ID NO:	Sequence
15			
	FL66	SEQ ID NO:30	5'CTATGCCATGGATAGATCGCTT-
			TCTACTTCC
	FL67	SEQ ID NO:31	5'CAAGCCCATGGAAACTTACAAG-
			GCTCAAAGA

20

#### Example 7

# Expression of Truncated Taf Polymerase

25 Mutein forms of the <u>Taf</u> polymerase lacking 5' to 3' exonuclease activity were constructed by introducing deletions in the 5'end of the <u>Taf</u> polymerase gene. Both 279 and 417 base pair deletions were created using the following protocol; an expression plasmid was 30 digested with restriction enzymes to excise the desired fragment, the fragment ends were repaired with Klenow and all four dNTP/s, to produce blunt ends, and the products were ligated to produce a new circular plasmid with the desired deletion. To express a 93 kilodalton, 35 5' to 3' exonuclease-deficient form of <u>Taf</u> polymerase, a 279 bp deletion comprising amino acids 2-93 was

generated. To express an 88 kilodalton, 5' to 3' exonuclease-deficient form of <u>Taf</u> polymerase, 417 bp deletion comprising amino acids 2-139 was generated.

To create a plasmid with codons 2-93 deleted, 5 pTaf03 was digested with NcoI and NdeI and the ends were repaired by Klenow treatment. The digested and repaired plasmid was diluted to 5 µg/ml and ligated under blunt end conditions. The dilute plasmid concentration favors intramolecular ligations. The 10 ligated plasmid was transformed into DG116. Mini-screen DNA preparations subjected were to restriction analysis and correct plasmids were confirmed by DNA sequence analysis. The resulting expression vector created by deleting a segment from 15 pTaf03 was designated pTaf09. A similar vector created from pTaf05 was designated pTaf10.

Expression vectors also were created with codons 2-139 deleted. The same protocol was used with the exception that the initial restriction digestion was 20 performed with <a href="NcoI">NcoI</a> and <a href="BglII">BglII</a>. The expression vector created from pTaf03 was designated pTaf11 and the expression vector created from pTaf05 was designated pTaf12.

25

30

# Example 8

Derivation and Expression of 5' to 3'
Exonuclease-Deficient, Thermostable DNA
Polymerase of Thermus species, Z05
Comprising Amino Acids 292 Through 834

To obtain a DNA fragment encoding a 5' to 3' exonuclease-deficient thermostable DNA polymerase from Thermus species Z05, a portion of the DNA polymerase gene comprising amino acids 292 through 834 is selectively amplified in a PCR with forward primer TZA292 and reverse primer TZR01 as follows:

5

20

25

50 pmoles TZA292
50 pmoles TZR01
10 ng <u>Thermus</u> sp. Z05 genomic DNA
2.5 units AmpliTag DNA polymerase
50 µM each dATP, dGTP, dCTP, dTTP

in an 80 µl solution containing 10 mM Tris-HCl pH 8.3, 50 mM KCl and overlaid with 100 µl of mineral oil. The reaction was initiated by addition of 20 µl containing 10 7.5 mM MgCl<sub>2</sub> after the tubes had been placed in an 80°C preheated cycler.

The genomic DNA was digested to completion with restriction endonuclease <u>Asp</u>718, denatured at 98°C for 5 minutes and cooled rapidly to 0°C. The sample was 15 cycled in a Perkin-Elmer Cetus Thermal Cycler according to the following profile:

STEP CYCLE to 96°C and hold for 20 seconds.

STEP CYCLE to 55°C and hold for 30 seconds.

RAMP to 72°C over 30 seconds and hold for 1 minute.

REPEAT profile for 3 cycles.

STEP CYCLE to 96°C and hold for 20 seconds.

STEP CYCLE to 65°C and hold for 2 minutes.

REPEAT profile for 25 cycles.

After last cycle HOLD for 5 minutes.

The intended 1.65 kb PCR product is purified by agarose gel elecctrophoresis, and recovered following 30 phenol-chloroform extraction and ethanol precipitation. The purified product is digested with restriction endonucleases NdeI and BqlII and ligated with NdeI/BamHI-digested and dephosphorylated plasmid vector pDG164 (U.S. Serial No. 455,967, filed December 22, 35 1989, Example 6B incorporated herein by reference). Ampicillin-resistant transformants of E. coli strain

DG116 are selected at 30°C and screened for the desired recombinant plasmid. Plasmid pZ05A292 encodes a 544 amino acid, 5' to 3' exonuclease-deficient Thermus sp. Z05 thermostable DNA polymerase analogous to the pLSG8 encoded protein of Example 2. The DNA polymerase activity is purified as in Example 2. The purified protein is deficient in 5' to 3' exonuclease activity, is more thermoresistant than the corresponding native enzyme and is particularly useful in PCR of G+C-rich templates.

	<u>Primer</u>	SEQ ID NO:	SEQUENCE
15	TZA292	SEQ ID NO:32	GTCGGCATATGGCTCCTGCTCCTTGAGGA- GGCCCCTGGCCCCCGCC
	TZR01	SEQ ID NO:33	GACGCAGATCTCAGCCCTTGGCGGAAAGCCA-GTCCTC
20			Example 9

Derivation and Expression of 5' to 3'
Exonuclease-Deficient, Thermostable DNA
Polymerase of <u>Thermus</u> species SPS17
Comprising Amino Acids 288 Through 830

To obtain a DNA fragment encoding 5' to 3' exonuclease-deficient thermostable DNA polymerase from Thermus species SPS17, a portion of the DNA polymerase gene comprising amino acids 288 through 830 is selectively amplified in a PCR with forward primer TSA288 and reverse primer TSR01 as follows:

-99-

50 pmoles TSA288

50 pmoles TSR01

10 ng Thermus sp. SPS17 genomic DNA

2.5 units AmpliTaq DNA polymerase

5 50 uM each dATP, dGTP, dCTP, dTTP

in an 80 µl solution containing 10 mM Tris-HCl pH 8.3, 50 mM KCl and overlaid with 100 µl of mineral oil. The reaction was initiated by addition of 20 µl containing 10 7.5 mM MgCl<sub>2</sub> after the tubes had been placed in an 80°C preheated cycler.

The genomic DNA was denatured at 98°C for 5 minutes and cooled rapidly to 0°C. The sample was cycled in a 15 Perkin-Elmer Cetus Thermal Cycler according to the following profile:

STEP CYCLE to 96°C and hold for 20 seconds. STEP CYCLE to 55°C and hold for 30 seconds.

20 RAMP to 72°C over 30 seconds and hold for 1 minute. REPEAT profile for 3 cycles.

STEP CYCLE to 96°C and hold for 20 seconds. STEP CYCLE to 65°C and hold for 2 minutes. REPEAT profile for 25 cycles.

After last cycle HOLD for 5 minutes.

25

The intended 1.65 kb PCR product is purified by agarose gel electrophoresis, and recovered following 30 phenol-chloroform extraction and ethanol precipitation. The purified product is digested with restriction endonucleases NdeI and BglII and ligated with NdeI/BamHI-digested and dephosphorylated plasmid vector pDG164 (U.S. Serial No. 455,967, filed December 12, 35 1989, Example 6B). Ampicillin- resistant transformants of E. coli strain DG116 are selected at 30°C and

screened for the desired recombinant plasmid. Plasmid pSPSA288 encodes a 544 amino acid, to exonuclease-deficient Thermus sp. SPS17 thermostable DNA polymerase analogous to the pLSG8 encoded protein 5 of Example 2. The DNA polymerase activity is purified as in Example 2. The purified protein is deficient in 5' to 3' exonuclease activity, is more thermoresistant corresponding the native enzyme and is particularly useful in PCR of G+C-rich templates.

10

	Primer	SEQ ID NO:	SEQUENCE
	TSA288	SEQ ID NO:34	GTCGGCATATGGCTCCTAAAGAAGCTGAGGA- GGCCCCCTGGCCCCCGCC
15	TSR01	SEQ ID NO:35	GACGCAGATCTCAGGCCTTGGCGGAAAGCCA-

### Example 10

GTCCTC

20

Derivation and Expression of 5' to 3'
Exonuclease-Deficient, Thermostable DNA
Polymerase of Thermus Thermophilus
Comprising Amino Acids 292 Through 834

25

35

To obtain a DNA fragment encoding a 5' to 3' exonuclease-deficient thermostable DNA polymerase from <a href="Thermus thermophilus">Thermus thermophilus</a>, a portion of the DNA polymerase gene comprising amino acids 292 through 834 is 30 selectively amplified in a PCR with forward primer TZA292 and reverse primer DG122 as follows;

50 pmoles TZA292
50 pmoles DG122
1 ng <u>Eco</u>RI digested plasmid pLSG22

2.5 units AmpliTaq DNA polymerase
50 µM each dATP, dGTP, dCTP, dTTP

in an 80 µl solution containing 10 mM Tris-HCl pH 8.3, 50 mM KCl and overlaid with 100 µl of mineral oil. The reaction was initiated by addition of 20 µl containing 7.5 mM MgCl<sub>2</sub> after the tubes had been placed in an 80°C preheated cycler.

Plasmid pLSG22 (U.S. Serial No. 455,967, filed December 22, 1989, Example 4A, incorporated herein by reference) was digested to completion with restriction 10 endonuclease <a href="EcoRI">EcoRI</a>, denatured at 98°C for 5 minutes and cooled rapidly to 0°C. The sample was cycled in a Perkin-Elmer Cetus Thermal Cycler according to the following profile:

- STEP CYCLE to 96°C and hold for 20 seconds.

  STEP CYCLE to 55°C and hold for 30 seconds.

  RAMP to 72°C over 30 seconds and hold for 1 minute.

  REPEAT profile for 3 cycles.
- 20 STEP CYCLE to 96°C and hold for 20 seconds.

  STEP CYCLE to 65°C and hold for 2 minutes.

  REPEAT profile for 20 cycles.

  After last cycle HOLD for 5 minutes.
- The intended 1.66 kb PCR product is purified by agarose gel electrophoresis, and recovered following phenol-chloroform extraction and ethanol precipitation. The purified product is digested with restriction BqlII and ligated endonucleases NdeI and 30 NdeI/BamHI-digested and dephosphorylated plasmid vector pDG164 (U.S. Serial No. 455,967, filed December 12, 1989, Example 6B). Ampicillin- resistant transformants of E. coli strain DG116 are selected at 30°C and screened for the desired recombinant plasmid. encodes a 544 amino acid, to 35 pTTHA292 exonuclease-deficient Thermus thermophilus thermostable

DNA polymerase analogous to the pLSG8 encoded protein of Example 2. The DNA polymerase activity is purified as in Example 2. The purified protein is deficient in 5' to 3' exonuclease activity, is more thermoresistant 5 than the corresponding native enzyme and is particularly useful in PCR of G+C-rich templates.

	<u>Primer</u>	SEO ID NO:	SEQUENCE
10	TZA292	SEQ ID NO:32	GTCGGCATATGGCTCCTGCTCCTTGAGGA- GGCCCCCTGGCCCCCGCC
1 5	DG122	SEQ ID NO:36	CCTCTAAACGGCAGATCTGATATCAACCCTT- GGCGGAAAGC

15

# Example 11

Derivation and Expression of 5' to 3'
Exonuclease-Deficient, Thermostable DNA
Polymerase of Thermosipho Africanus
Comprising Amino Acids 285 Through 892

To obtain a DNA fragment encoding a 5' to 3' exonuclease-deficient thermostable DNA polymerase from 25 Thermosipho africanus, a portion of the DNA polymerase gene comprising amino acids 285 through 892 is selectively amplified in a PCR with forward primer TAFI285 and reverse primer TAFRO1 as follows:

- 30 50 pmoles TAFI285
  - 50 pmoles TAFR01
  - 1 ng plasmid pBSM:TafRV3' DNA
  - 2.5 units AmpliTaq DNA polymerase
  - 50 μM each dATP, dGTP, dCTP, dTTP

35

in an 80  $\mu$ l solution containing 10 mM Tris-HCl pH 8.3, 50 mM KCl and overlaid with 100  $\mu$ l of mineral oil. The

20

reaction was initiated by addition of 20  $\mu l$  containing 7.5 mM MgCl $_2$  after the tubes had been placed in an 80°C preheated cycler.

Plasmid pBSM:TafRV'3 (obtained as described in CETUS CASE 2583.1, EX 4, p53, incorporated herein by reference) was digested with <a href="EcoRI">EcoRI</a> to completion and the DNA was denatured at 98°C for 5 minutes and cooled rapidly to 0°C. The sample was cycled in a Perkin-Elmer Cetus Thermal Cycler according to the following profile:

STEP CYCLE to 95°C and hold for 30 seconds. STEP CYCLE to 55°C and hold for 30 seconds.

RAMP to 72°C over 30 seconds and hold for 1 minute.
REPEAT profile for 3 cycles.

STEP CYCLE to 95°C and hold for 30 minutes. STEP CYCLE to 65°C and hold for 2 minutes. REPEAT profile for 20 cycles. After last cycle HOLD for 5 minutes.

The intended 1.86 kb PCR product is purified by agarose gel electrophoresis, and recovered following 25 phenol-chloroform extraction and ethanol precipitation. The purified product is digested with restriction and ligated and BamHI endonucleases NdeI NdeI/BamHI-digested and dephosphorylated plasmid vector pDG164 (U.S. Serial No. 455,967, filed December 22, 30 1989, Example 6B). Ampicillin- resistant transformants of E. coli strain DG116 are selected at 30°C and screened for the desired recombinant plasmid. acid, to amino pTAF1285 encodes а 609 africanus exonuclease-deficient Thermosipho polymerase the analogous to DNA 35 thermostable pTMA15-encoded protein of Example The DNA 3.

polymerase activity is purified as in Example 3. The purified protein is deficient in 5' to 3' exonuclease activity, is more thermoresistant than the corresponding native enzyme and is particularly useful 5 in PCR of G+C-rich templates.

Primer SEO ID NO: SEQUENCE

appended claims.

TAFI285 SEQ ID NO:37 GTCGGCATATGATTAAAGAACTTAATTTACAAGAAAAATTAGAAAAGG

TAFR01 SEQ ID NO:38 CCTTTACCCCAGGATCCTCATTCCCACTCTTTTCCATAATAAACAT

15 The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the cell lines deposited, since the deposited embodiment is intended as a single 20 illustration of one aspect of the invention and any cell lines that are functionally equivalent are within the scope of this invention. The deposits of materials therein does not constitute an admission that the written description herein contained is inadequate to 25 enable the practice of any aspect of the invention, including the best mode thereof, nor are the deposits to be construed as limiting the scope of the claims to specific illustrations that they represent. Indeed, various modifications of the invention 30 addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the

### SEQUENCE LISTING

- (1) GENERAL INFORMATION:
  - (i) APPLICANT: Gelfand, David H.
    Abramson, Richard D.
  - (ii) TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF THERMOSTABLE DNA POLYMERASES
  - (iii) NUMBER OF SEQUENCES: 38
  - (iv) CORRESPONDENCE ADDRESS:
    - (A) ADDRESSEE: Cetus Corporation
    - (B) STREET: 1400 Fifty-third Street
    - (C) CITY: Emeryville
    - (D) STATE: California
    - (F) ZIP: 94608
  - (v) COMPUTER READABLE FORM:
    - (A) MEDIUM TYPE: Floppy disk
    - (B) COMPUTER: IBM PC compatible
    - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
    - (D) SOFTWARE: WordPerfect 5.0
  - (vi) CURRENT APPLICATION DATA:
    - (A) APPLICATION NUMBER: WO
    - (B) FILING DATE:
    - (C) CLASSIFICATION:
  - (vii) PRIOR APPLICATION DATA:
    - (A) APPLICATION NUMBER: US 590,490
    - (B) FILING DATE: 28-SEP-1990
  - (vii) PRIOR APPLICATION DATA:
    - (A) APPLICATION NUMBER: US 590,466
    - (B) FILING DATE: 28-SEP-1990
  - (vii) PRIOR APPLICATION DATA:
    - (A) APPLICATION NUMBER: US 590,213
    - (B) FILING DATE: 28-SEP-1990
  - (vii) PRIOR APPLICATION DATA:
    - (A) APPLICATION NUMBER: US 523,394
    - (B) FILING DATE: 15-MAY-1990
  - (vii) PRIOR APPLICATION DATA:
    - (A) APPLICATION NUMBER: US 143,441
    - (B) FILING DATE: 12-JAN-1988
  - (vii) PRIOR APPLICATION DATA:
    - (A) APPLICATION NUMBER: US 063,509
    - (B) FILING DATE: 17-JUN-1987

- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 899,241
  - (B) FILING DATE: 22-AUG-1986
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 746,121
  - (B) FILING DATE: 15-AUG-1991
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: WO PCT/US90/07641
  - (B) FILING DATE: 21-DEC-1990
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 585,471
  - (B) FILING DATE: 20-SEP-1990
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 455,611
  - (B) FILING DATE: 22-DEC-1989
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 609,157
  - (B) FILING DATE: 02-NOV-1990
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 557,517
  - (B) FILING DATE: 24-JUL-1990
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Sias Ph.D, Stacey R.
  - (B) REGISTRATION NUMBER: 32,630
  - (C) REFERENCE/DOCKET NUMBER: Case No. 2580
  - (ix) TELECOMMUNICATION INFORMATION:
    - (A) TELEPHONE: 415-420-3300
- (2) INFORMATION FOR SEQ ID NO:1:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2499 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (iii) HYPOTHETICAL: NO
  - (iv) ANTI-SENSE: NO
  - (vi) ORIGINAL SOURCE:
    - (A) ORGANISM: Thermus aquaticus

# (ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..2496

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG	AGG	GGG	ATG	CTG	CCC	CTC	TTT	GAG	CCC	AAG	GGC	CGG	GTC	CTC	CTG	;	-48
Met 1	Arg	Gly	Met	Leu 5	Pro	Leu	Phe	Glu	Pro 10		s Gl	y Ar	g Va		u Le 5	u	
GTG	GAC	GGC	CAC	CAC	CTG	GCC	TAC	CGC	ACC	TTC	CAC	GCC	CTG	AAG	GGC	:	96
Val	Asp	Gly	His 20		Leu	Ala	Tyr	Arg 25		Phe	e Hi	s Ala	a Le		s Gl	У	
CTC	ACC	ACC	AGC	CGG	GGG	GAG	CCG	GTG	CAG	GCG	GTC	TAC	GGC	TTC	GCC		144
Leu	Thr	Thr 35		Arg	Gly	Glu	Pro 40		Gln	Ala	ı Va	1 Ty:		y Ph	e Al	a	
AAG	AGC	CTC	CTC	AAG	GCC	CTC	AAG	GAG	GAC	GGG	GAC	GCG	GTG	ATC	GTG		192
Lys	Ser 50	Leu	Leu	Lys	Ala	Leu 55	Lys	Glu	Asp	G1y	Ası 60	Ala	a Vai	1 11	e Va	1	
GTC	TTT	GAC	GCC	AAG	GCC	CCC	TCC	TTC	CGC	CAC	GAG	GCC	TAC	GGG	GGG		240
Val 65	Phe	Asp	Ala	Lys	Ala 70	Pro	Ser	Phe	Arg	His 75		ı Ala	Туз	r Gly	y Gl <sub>3</sub>		
TAC	AAG	GCG	GGC	CGG	GCC	ccc	ACG	CCG	GAG	GAC	TTT	ccc	CGG	CAA	CTC		288
Tyr	Lys	Ala	Gly	Arg 85	Ala	Pro	Thr	Pro	Glu 90	Asp	Phe	Pro	Arg	g G1r 95		ı	
GCC	CTC	ATC	AAG	GAG	CTG	GTG	GAC	CTC	CTG	GGG	CTG	GCG	CGC	CTC	GAG		336
Ala	Leu	Ile	Lys 100	Glu	Leu	Val	Asp	Leu 105	Leu	Gly	Leu	Ala	Arg 110		ı Glu	1	
GTC	CĊG	GGC	TAC	GAG	GCG	GAC	GAC	GTC	CTG	GCC	AGC	CTG	GCC	AAG	AAG	:	384
Val	Pro	Gly 115	Tyr	Glu	Ala	Asp	Asp 120	Val	Leu	Ala	Ser	Leu 125		Lys	Lys	5	
GCG	GAA	AAG	GAG	GGC	TAC	GAG	GTC	CGC .	ATC	CŤC	ACC	GCC	GAC	AAA	GAC	4	432
	Glu 130	Lys	Glu	Gly	Tyr	<b>Glu</b> 135	Val	Arg	Ile	Leu	Thr 140		Asp	Lys	Asp	)	
CTT '	TAC	CAG	CTC	CTT	TCC (	GAC	CGC	ATC (	CAC	GTC	CTC	CAC	ссс	GAG	GGG	4	480
Leu '	Tyr	Gln	Leu	Leu	Ser 150	Asp	Arg	Ile	His	Val 155	Leu	His	Pro	Glu	Gly 160		

TAC	CTC	ATC	ACC	CCG	GCC	TGG	CTT	TGG	GAA	AAG	TAC	GGC	CTG	AGG	CCC	528	
Tyr	Leu	Ile	Thr	Pro 165		Trp	Leu	Trp	Glu 170		з Туг	Gly	y Let	a Arį 17	g Pro		
GAC	CAG	TGG	GCC	GAC	TAC	CGG	GCC	CTG	ACC	GGG	GAC	GAG	TCC	GAC	AAC	576	
Asp	Gln	Trp	Ala 180	Asp	Tyr	Arg	Ala	Leu 185		Gly	Asp	Glu	1 Sei 190		Asn		
CTT	CCC	GGG	GTC	AAG	GGC	ATC	GGG	GAG	AAG	ACG	GCG	AGG	AAG	CTT	CTG	624	
Leu	Pro	Gly 195	Val	Lys	Gly	Ile	Gly 200		Lys	Thr	Ala	Arg 205		Leu	Leu		
GAG	GAG	TGG	GGG	AGC	CTG	GAA	GCC	СТС	CTC	AAG	AAC	CTG	GAC	CGG	CTG	672	
Glu	Glu 210	Trp	Gly	Ser	Leu	Glu 215	Ala	Leu	Leu	Lys	Asn 220		Asp	Arg	; Leu		
AAG	CCC	GCC	ATC	CGG	GAG	AAG	ATC	CTG	GCC	CAC	ATG	GAC	GAT	CTG	AAG	720	
Lys 225	Pro	Ala	Ile	Arg	Glu 230	Lys	Ile	Leu	Ala	His 235	Met	Asp	Asp	Leu	Lys 240		
CTC	TCC	TGG	GAC	CTG	GCC	AAG	GTG	CGC	ACC	GAC	CTG	ссс	CTG	GAG	GTG	768	
Leu	Ser	Trp	Asp	Leu 245	Ala	Lys	Val	Arg	Thr 250	Asp	Leu	Pro	Leu	Glu 255	Val		
GAC	TTC	GCC	AAA	AGG	CGG	GAG	CCC -	GAC	CGG	GAG	AGG	CTT .	AGG	GCC	TTT	816	
Asp	Phe	Ala	Lys 260	Arg	Arg	Glu	Pro	Asp 265	Arg	Glu	Arg	Leu	Arg 270	Ala	Phe		
CTG	GAG	AGG	CTT	GAG	TTT	GGC .	AGC	CTC	CTC	CAC	GAG '	TTC	GGC (	CTT (	CTG	864	
Leu	Glu	Arg 275	Leu	Gľu	Phe	Gly	Ser 280	Leu	Leu	His	Glu	Phe 285	Gly	Leu	Leu .		
GAA	AGC	CCC	AAG (	GCC	CTG	GAG (	GAG (	GCC (	ccc :	TGG (	ccc (	CCG (	CCG (	GAA (	GGG <sub>.</sub>	912	
Glu	Ser 290	Pro	Lys	Ala	Leu	Glu 295	G1u	Ala	Pro	Trp	Pro 300	Pro	Pro	Glu	Gly		
GCC	TTC	GTG	GGC 1	TTT (	GTG ·	CTT :	rcc (	CGC A	AAG (	GAG (	CCC A	ATG T	rgg (	GCC (	SAT	960	
Ala 305	Phe	Val	Gly	Phe	Val 310	Leu	Ser	Arg	Lys	Glu 315	Pro	Met	Trp	Ala	Asp 320		
CTT	CTG	GCC	CTG (	GCC (	GCC (	GCC A	AGG (	GGG (	GGC (	CGG (	GTC (	CAC (	ccc c	GCC C	cc	1008	
Leu	Leu .	Ala	Leu .	Ala . 325	Ala	Ala	Arg		Gly 330	Arg	Val	His	Arg	Ala	Pro		

GAG	CCT	TAT	AAA	GCC	CTC	AGG	GAC	CTG	AAG	GAG	GCG	CGG	GGG	CTT	CTC	1056
Glu	Pro	Tyr	Lys 340		Leu	Arg	Asp	Leu 345		Glu	ı Ala	a Arg	350		ı Leu	
GCC	AAA	GAC	CTG	AGC	GTT	CTG	GCC	CTG	AGG	GAA	GGC	CTT	GGC	CTC	CCG	1104
Ala	Lys	Asp 355		Ser	Val	Leu	Ala 360		Arg	Glu	Gly	7 Leu 365		.Leu	Pro	
CCC	GGC	GAC	GAC	CCC	ATG	CTC	CTC	GCC	TAC	CTC	CTG	GAC	CCT	TCC	AAC	1152
Pro	Gly 370	Asp	Asp	Pro	Met	Leu 375		Ala	Tyr	Leu	1 Leu 380		Pro	Ser	Asn	
ACC	ACC	ccc	GAG	GGG	GTG	GCC	CGG	CGC	TAC	GGC	GGG	GAG	TGG	ACG	GAG	1200
Thr 385	Thr	Pro	Glu	Gly	Val 390	Ala	Arg	Arg	Tyr	Gly 395		Glu	Trp	Thr	Glu 400	
GAG	GCG	GGG	GAG	CGG	GCC	GCC	CTT	TCC	GAG	AGG	CTC	TTC	GCC .	AAC	CTG	1248
Glu	Ala	Gly	Glu	Arg 405	Ala	Ala	Leu	Ser	G1u 410	Arg	Leu	Phe	Ala	Asn 415	Leu	
TGG	GGG	AGG	CTT	GAG	GGG	GAG	GAG	AGG	CTC	CTT	TGG	CTT	TAC	CGG	GAG	1296
Trp	Gly	Arg	Leu 420	Glu	Gly	Glu	Glu	Arg 425	Leu	Leu	Trp	Leu	Tyr 430	Arg	Glu	
GTG	GAG	AGG	CCC	CTT	TCC	GCT	GTC	CTG	GCC	CAC	ATG	GAG	GCC A	ACG	GGG	1344
Val	Glu	Arg 435	Pro	Leu	Ser	Ala	Val 440	Leu	Ala	His	Met	Glu 445	Ala	Thr	Gly	
GTG	CGC	CTG	GAC	GTG	GCC	TAT	CTC	AGG	GCC	TTG	TCC	CTG	GAG (	GTG	GCC	1392
	Arg 450	Leu	Asp	Val	Ala	Tyr 455	Leu	Arg	Ala	Leu	Ser 460	Leu	Glu	Val	Ala	
GAG	GAG	ATC	GCC	CGC	CTC	GAG	GCC	GAG	GTC .	TTC	CGC	CTG	GCC (	GGC (	CAC	1440
Glu 465	Glu	Ile	Ala	Arg	Leu 470	Glu	Ala	G1u	Val	Phe 475	Arg	Leu	Ala	Gly	His 480	
CCC	TTC	AAC	CTC	AAC	TCC	CGG	GAC	CAG	CTG (	GAA .	AGG	GTC (	CTC 1	TT (	GAC	1488
Pro	Phe	Asn	Leu	Asn 485	Ser	Arg	Asp	Gln	Leu 490	Glu	Arg	Val	Leu	Phe 495	Asp	
GAG	CTA	GGG	CTT	ccc	GCC A	ATC	GGC .	AAG .	ACG (	GAG A	AAG A	ACC (	GGC A	AG (	CGC	1536
Glu	Leu	Gly	Leu 500	Pro	Ala	Ile	Gly	Lys 505	Thr	Glu	Lys	Thr	Gly 510	Lys	Arg	

TCC	ACC	AGC	GCC	GCC	GTC	CTG	GAG	GCC	CTC	CGC	GAG	GCC	CAC	CCC	ATC	1584
Ser	Thr	Ser 515		Ala	Val	Leu	G1: 520		Leu	ı Ar	g Glu	a Ala 52		s Pr	o Ile	
GTG	GAG	AAG	ATC	CTG	CAG	TAC	CGG	GAG	CTC	ACC	AAG	CTG	AAG	AGC	ACC	1632
Val	Glu 530	Lys	Ile	Leu	Gln	Tyr 535	Arg	Glu	Leu	Thi	r Lys 540		ı Ly:	s Se	r Thr	
TAC	ATT	GAC	CCĆ	TTG	CCG	GAC	CTC	ATC	CAC	CCC	AGG	ACG	GGC	CGC	CTC	1680
Tyr 545	Ile	Asp	Pro	Leu	Pro 550	Asp	Leu	lle	His	Pro 555		g Thi	c Gly	y Ar	g Leu 560	
CAC	ACC	CGC	TTC	AAC	CAG	ACG	GCC	ACG	GCC	ACG	GGC	AGG	CTA	AGT	AGC	1728
His	Thr	Arg	Phe	Asn 565	Gln	Thr	Ala	Thr	Ala 570		Gly	Arg	g Let	sei 575	r Ser	
TCC	GAT	CCC	AAC	CTC	CAG	AAC	ATC	CCC	GTC	CGC	ACC	CCG	CTT	GGG	CAG	1776
Ser	Asp	Pro	Asn 580	Leu	Gln	Asn	Ile	Pro 585	Val	Arg	Thr	Pro	Let 590	_	/ Gln	
AGG	ATC	CGC	CGG	GCC	TTC	ATC	GCC	GAG	GAG	GGG	TGG	CTA	TTG	GTG	GCC	1824
Arg	Ile	Arg 595	Arg	Ala	Phe	Ile	Ala 600	Glu	Glu	Gly	Trp	Leu 605		ı Val	Ala	
CTG	GAC	TAT	AGC	CAG	ATA	GAG	CTC-	AGG	GTG	CTG	GCC	CAC	CTC	TCC	GGC	1872
Leu	Asp 610	Tyr	Ser	Gln	Ile	Glu 615	Leu	Arg	Val	Leu	Ala 620	His	Leu	Ser	Gly	
GAC	GAG	AAC	CTG	ATC	CGG	GTC	TTC	CAG	GAG	GGG	CGG	GAC	ATC	CAC	ACG	1920
Asp 625	G1u	Asn	Leu	Ile	Arg 630	Val	Phe	Gln	Glu	Gly 635		Asp	Ile	His	Thr 640	
GAG	ACC	GCC	AGC	TGG	ATG	TTC	GGC	GTC	ccc	CGG	GAG	GCC	GTG	GAC	ссс	1968
Glu	Thr	Ala	Ser	Trp 645	Met	Phe	Gly	Val	Pro 650	Arg	Glu	Ala	Val	Asp 655	Pro	
CTG	ATG	CGC	CGG	GCG	GCC .	AAG A	ACC	ATC A	AAC '	TTC	GGG (	GTC	CTC	TAC	GGC	2016
Leu	Met	Arg	Arg 660	Ala	Ala	Lys	Thr	11e 665	Asn	Phe	Gly	Val	Leu 670	-	Gly	
ATG <sup>*</sup>	TCG	GCC	CAC	CGC	CTC :	rcc (	CAG	GAG (	CTA (	GCC .	ATC (	CCT '	TAC	GAG	GAG	2064
Met	Ser	Ala 675	His	Arg	Leu		Gln 680	Glu	Leu	Ala	Ile	Pro 685	Tyr	Glu	Glu	

GCC	CAG	GCC	TTC	ATT	GAG	CGC	TAC	TTT	CAG	AGC	TTC	CCC	AAG	GTG	CGG	3	2112
Ala	Gln 690	Ala	Phe	Ile	Glu	Arg 695		Phe	Gln	Ser	700	Pro	Lys	Val	. Ar	g	
GCC	TGG	ATT	GAG	AAG	ACC	CTG	GAG	GAG	GGC	AGG	AGG	CGG	GGG	TAC	GTG	;	2160
Ala 705	Trp	Ile	Glu	Lys	Thr 710	Leu	Glu	Glu	Gly	Arg 715		, Arg	Gly	Tyr	72	1 20	
GAG	ACC	CTC	TTC	GCC	CGC	CGC	CGC	TAC	GTG	CCA	GAC	CTA	GAG	GCC	CGG	}	2208
Glu	Thr	Leu	Phe	Gly 725	Arg	Arg	Arg	Tyr	Val 730	Pro	Asp	Leu	Glu	735	Ar	g	
GTG	AAG	AGC	GTG	CGG	GAG	GCG	GCC	GAG	CGC	ATG	GCC	TTC	AAC	ATG	CCC	:	2256
Val	Lys	Ser	Val 740	Arg	Glu	Ala	Ala	Glu 745		Met	Ala	Phe	Asn 750	ifet	Pr	0	
GTC	CAG	GGC	ACC	GCC	GCC	GAC	CTC	ATG	AAG	CTG	GCT	ATG	GTG	AAG	CTC		2304
Val	Gln	G1y 755	Thr	Ala	Ala	Asp	Leu 760		Lys	Leu	Ala	Met 765		Lys	Le	u	
гтс	CCC	AGG	CTG	GAG	GAA	ATG	GGG	GCC	AGG	ATG	CTC	CTT	CAG	GTC	CAC		2352
Phe	Pro 770	Arg	Leu	Glu	Glu	Met 775	Gly	Ala	Arg	Met	Leu 780	Leu	Gln	Val	Hi	s	
GAC	GAG	CTG	CTC	CTC	GAG	GCC	CCA	AAA	GAG	AGG	GCG	GAG	GCC	GTG	GCC		2400
Asp 785	Glu	Leu	Val	Leu	Glu 790	Ala	Pro	Lys	Glu	Arg 795	Ala	Glu	Ala	Val	A1 80	a 0	
CGG	CTG	GCC	AAG	GAG	GTC	ATG	GAG	GGG	GTG	TAT	CCC	CTG	GCC	GTG	CCC		2448
Arg	Leu	Ala	Lys	G1u 805	Val	Met	Glu	Gly	Val 810	Tyr	Pro	Leu	Ala	Val 815	Pr	o	
CTG	GAG	GTG	GAG	GTG	GGG	ATA	GGG	GAG	GAC	TGG	CTC	TCC	GCC .	AAG (	GAG		2496
Leu	Glu	Val	Glu 820	Val	Gly	Ile	Gly	Glu 825	Asp	Trp	Leu	Ser	Ala 830	Lys	Gl	u	
[GA																	2499

# (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 832 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu 1 5 10 15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly
20 25 30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala 35 40 45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile Val 50 55 60

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly Gly 65 70 75 80

Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu 85 90 95

Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu Glu 100 105 110

Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys Lys 115 120 125

Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys Asp 130 135 140

Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu Gly 145 150 155 160

Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro 165 170 175

Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp Asn 180 185 190

Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu Leu 195 200 205

Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg Leu 210 215 220

Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu Lys 225 230 235 240

Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val
245 250 255

Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe 260 265 270

Lev	ı Glu	1 Arg 275		ı Glu	ı Phe	Gly	Ser 280		ı Lev	His	: Glu	Phe 285		Leu	Leu
Glu	Ser 290		Lys	Ala	Leu	Glu 295		Ala	Pro	Trp	9 Pro		Pro	Glu	Gly
Ala 305		Val	. Gly	Phe	Val 310		Ser	Arg	, Lys	Glu 315		Met	Trp	Ala	Asp 320
Leu	Leu	Ala	Leu	Ala 325		Ala	Arg	Gly	Gly 330		Val	His	Arg	Ala 335	
Glu	Pro	Tyr	Lys 340		Leu	Arg	Asp	Leu 345		Glu	Ala	Arg	Gly 350	Leu	Leu
Ala	Lys	Asp 355		Ser	Val	Leu	Ala 360	Leu	Arg	Glu	Gly	Leu 365	Gly	Leu	Pro
Pro	Gly 370		Asp	Pro	Met	Leu 375		Ala	Tyr	Leu	Leu 380	Asp	Pro	Ser	Asn
Thr 385		Pro	Glu	Gly	Val 390	Ala	Arg	Arg	Tyr	Gly 395	Gly	Glu	Trp	Thr	G1u 400
Glu	Ala	Gly	Glu	Arg 405	Ala	Ala	Leu	Ser	Glu 410	Arg	Leu	Phe	Ala	Asn 415	Leu
Trp	Gly	Arg	Leu 420	Glu	Gly	Glu	Glu	Arg 425	Leu	Leu	Trp	Leu	Tyr 430	Arg	Glu
Val	Glu	Arg 435	Pro	Leu	Ser	Ala	Val 440	Leu	Ala	His	Met	Glu 445	Ala	Thr	Gly
Val	Arg 450	Leu	Asp	Val	Ala	Tyr 455	Leu	Arg	Ala	Leu	Ser 460	Leu	Glu	Val	Ala
Glu 465	Glu	Ile	Ala	Arg	Leu 470	Glu	Ala	Glu	Val	Phe 475	Arg	Leu	Ala	Cly	His 480
Pro	Phe	Asn	Leu	Asn 485	Ser	Arg	Asp	Gln	Leu 490	Glu	Arg	Val	Leu	Phe ∴95	Asp
Glu	Leu	Gly	Leu 500	Pro	Ala	Ile	Gly	Lys 505	Thr	Glu	Lys	Thr	Gly 510	!.ys	Arg
Ser	Thr	Ser 515	Ala	Ala	Val	Leu	Glu 520	Ala	Leu	Arg	Glu	Ala 525	His	Pro	Ile
Val	G1u 530	Lys	Ile	Leu	Gln	Tyr 535	Arg	Glu	Leu	Thr	Lys 540	Leu	Lys	Ser	Thr
Tyr 545	Ile	Asp	Pro	Leu	Pro 550	Asp	Leu	Ile	His	Pro 555	Arg	Thr	G1y	Arg	Leu 560
His	Thr	Arg	Phe	Asn 565	Gln	Thr	Ala	Thr	Ala 570	Thr	Gly	Arg	Leu	Ser 575	Ser

Ser	Asp	Pro	Asn 580	Leu	Gln	Asn	Ile	Pro 585	Val	Arg	Thr	Pro	Leu 590	Gly	Gln
Arg	Ile	Arg 595	Arg	Ala	Phe	Ile	Ala 600	Glu	Glu	Gly	Trp	Leu 605	Leu	Val	Ala
Leu	Asp 610	Tyr	Ser	Gln	Ile	Glu 615	Leu	Arg	Val	Leu	Ala 620	His	Leu	Ser	Gly
Asp 625	Glu	Asn	Leu	Ile	Arg 630	Val	Phe	Gln	Glu	G1y 635	Arg	Asp	Ile	His	Thr 640
Glu	Thr	Ala	Ser	Trp 645	Met	Phe	Gly	Val	Pro 650	Arg	Glu	Ala	Val	Asp 655	Pro
Leu	Met	Arg	Arg 660	Ala	Ala	Lys	Thr	Ile 665	Asn	Phe	Gly	Val	Leu 670	Tyr	Gly
Met	Ser	Ala 675	His	Arg	Leu	Ser	Gln 680	Glu	Leu	Ala	Ile	Pro 685	Tyr	Glu	Glu
Ala	Gln 690	Ala	Phe	Ile	Glu	Arg 695	Tyr	Phe	Gln	Ser	Phe 700	Pro	Lys	Val	Arg
Ala 705	Trp	Ile	Glu	Lys	Thr 710	Leu	Glu	Glu	Gly	Arg 715	Arg	Arg.	Gly	Tyr	Val 720
Glu	Thr	Leu	Phe	G1y 725	Arg	Arg	Arg	Tyr	Val 730	Pro	Asp	Leu	Glu	Ala 735	Arg
Val	Lys	Ser	Val 740	Arg	Glu	Ala	Ala	Glu 745	Arg	Met	Ala	Phe	Asn 750	Met	Pro
Val	Gln	Gly 755	Thr	Ala	Ala	Asp	Leu 760	Met	Lys	Leu	Ala	Met 765	Val	Lys	Leu
Phe	Pro 770	Arg	Leu	Glu	Glu	Met 775	Gly	Ala	Arg	Met	Leu 780	Leu	G1n	Val	His
Asp 785	Glu	Leu	Val	Leu	Glu 790	Ala	Pro	Lys	Glu	Arg 795	Ala	Glu	Ala	Val	Ala 800
Arg	Leu	Ala	Lys	Glu 805	Val	Met	Glu	Gly	Val 810	Tyr	Pro	Leu	Ala	Val 815	Pro

Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu 820 825 830

## (2) INFORMATION FOR SEQ ID NO:3:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2682 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

í	(D)	TOPOL	OGY:	1 i	near
1		TOTOL	~~.	**	11007

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Thermotoga maritima

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..2679

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATG GCG AGA CTA TTT CTC TTT GAT GGA ACT GCT CTG GCC TAC AGA GCG 48 Met Ala Arg Leu Phe Leu Phe Asp Gly Thr Ala Leu Ala Tyr Arg Ala TAC TAT GCG CTC GAT AGA TCG CTT TCT ACT TCC ACC GGC ATT CCC ACA 96 Tyr Tyr Ala Leu Asp Arg Ser Leu Ser Thr Ser Thr Gly Ile Pro Thr AAC GCC ACA TAC GGT GTG GCG AGG ATG CTG GTG AGA TTC ATC AAA GAC 144 Asn Ala Thr Tyr Gly Val Ala Arg Met Leu Val Arg Phe Ile Lys Asp CAT ATC ATT GTC GGA AAA GAC TAC GTT GCT GTG GCT TTC GAC AAA AAA 192 His Ile Ile Val Gly Lys Asp Tyr Val Ala Val Ala Phe Asp Lys Lys GCT GCC ACC TTC AGA CAC AAG CTC CTC GAG ACT TAC AAG GCT CAA AGA 240 Ala Ala Thr Phe Arg His Lys Leu Leu Glu Thr Tyr Lys Ala Gln Arg 288 CCA AAG ACT CCG GAT CTC CTG ATT CAG CAG CTT CCG TAC ATA AAG AAG Pro Lys Thr Pro Asp Leu Leu Ile Gln Gln Leu Pro Tyr Ile Lys Lys CTG GTC GAA GCC CTT GGA ATG AAA GTG CTG GAG GTA GAA GGA TAC GAA 336 Leu Val Glu Ala Leu Gly Met Lys Val Leu Glu Val Glu Gly Tyr Glu

GCG	GAC	GAT	ATA	TTA	GCC	ACT	CTG	GCT	GTG	AAG	GGG	CTT	CCG	CTT	TTT	384
Ala	Asp	Asp 115	Ile	Ile	Ala	Thr	Leu 120		Val	Lys	Gly	Leu 125		Leu	ı Phe	3
GAT	GAA	ATA	TTC	ATA	GTG	ACC	GGA	GAT	AAA	GAC	ATG	CTT	CAG	CTT	GTG	432
Asp	Glu 130	Ile	Phe	Ile	Val	Thr 135	Gly	Asp	Lys	Asp	Met 140		Gln	Leu	ı Val	Ĺ
AAC	GAA	AAG	ATC	AAG	GTG	TGG	CGA	ATC	GTA	AAA	GGG	ATA	TCC	GAT	CTG	480
Asn 145	Glu	Lys	Ile	Lys	Val 150	Trp	Arg	Ile	Val	Lys 155		Ile	Ser	Ası	Let 160	
GAA	CTT	TAC	GAT	GCG	CAG	AAG	GTG	AAG	GAA	AAA	TAC	GGT	GTT	GAA	CCC	528
Glu	Leu	Tyr	Asp	Ala 165	Gln	Lys	Val	Lys	Glu 170	_	Tyr	Gly	Val	G1u 175		>
CAG	CAG	ATC	CCG	GAT	CTT	CTG	GCT	CTA	ACC	GGA	GAT	GAA	ATA	GAC	AAC	576
Gln	Gln	Ile	Pro 180	Asp	Leu	Leu	Ala	Leu 185		Gly	Asp	Glu	Ile 190	_	Ası	n
ATC	CCC	GGT	GTA	ACT	GGG	ATA	GGT	GAA	AAG	ACT	GCT	GTT	CAG	CTT	CTA	624
Ile	Pro	Gly 195	Val	Thr	Gly	Ile	Gly 200		Lys	Thr	Ala	Val 205		Let	ı Lev	1
GAG	AAG	TAC	AAA	GAC	CTC	GAA	GAC	ATA	CTG	TAA	CAT	GTT	CGC	GAA	CTT	672
Glu	Lys 210	Tyr	Lys	Asp	Leu	Glu 215		Ile	Leu	. Asn	His 220		Arg	G1	ı Leı	a.
CCT	CAA	AAG	GTG	AGA	AAA	GCC	CTG	CTT	CGA	GAC	AGA	GAA	AAC	GCC	ATT	720
Pro 225	Gln	Lys	Val	Arg	Lys 230	Ala	Leu	Leu	Arg	Asp 235	_	, Glu	Asn	Ala	11e 240	
CTC	AGC	AAA	AAG	CTG	GCG	ATT	CTG	GAA	ACA	AAC	GTT	CCC	ATT	GAA	ATA	768
Leu	Ser	Lys	Lys	Leu 245		Ile	Leu	Glu	Thr 250		Val	Pro	Ile	G1: 255		е
AAC	TGG	GAA	GAA	CTT	CGC	TAC	CAG	GGC	TAC	GAC	AGA	GAG	AAA	CTC	TTA	816
Asn	Trp	Glu	Glu 260	Leu	Arg	Tyr	Gln	Gly 265	_	Asp	Arg	; Glu	Lys 270		ı Lei	ı
CCA	CTT	TTG	AAA	GAA	CTG	GAA	TTC	GCA	TCC	ATC	ATG	AAG	GAA	CTT	CAA	864
Pro	Leu	Leu 275	Lys	G1u	Leu	Glu	Phe 280		Ser	Ile	Met	Lys 285		ı 1.ev	ı Glı	n

CTG	TAC	GAA	GAG	TCC	GAA	CCC	GTT	GGA	TAC	AGA	ATA	GTG	AAA	GAC	CTA	912
Leu	Tyr 290	Glu	Glu	Ser	Glu	Pro 295		Gly	Tyr	Arg	11e 300		Lys	: Ası	Leu	
GTG	GAA	TTT	GAA	AAA	CTC	ATA	GAG	AAA	CTG	AGA	GAA	TCC	CCT	TCG	TTC	960
Val 305	Glu	Phe	Glu	Lys	Leu 310	Ile	Glu	Lys	Leu	Arg 315		Ser	Pro	Sei	Phe 320	
GCC	ATA	GAT	CTT	GAG	ACG	TCT	TCC	CTC	GAT	CCT	TTC	GAC	TGC	GAC	ATT	1008
Ala	Ile	Asp	Leu	Glu 325	Thr	Ser	Ser	Leu	Asp 330		Phe	Asp	Cys	335	Ile	
GTC	GGT	ATC	TCT	GTG	TCT	TTC	AAA	CCA	AAG	GAA	GCG	TAC	TAC	ATA	CCA	1056
Val	Gly	Ile	Ser 340	Val	Ser	Phe	Lys	Pro 345	Lys	Glu	Ala	Tyr	Tyr 350	Ile	Pro	
CTC	CAT	CAT	AGA	AAC	GCC	CAG	AAC	CTG	GAC	GAA	AAA	GAG	GTT	CTG	AAA	1104
Leu	His	His 355	Arg	Asn	Ala	Gln	Asn 360		Asp	Glu	Lys	G1u 365	Val	Let	Lys	
AAG	CTC	AAA	GAA	ATT	CTG	GAG	GAC	CCC	GGA	GCA	AAG	ATC	GTT	GCT	CAG	1152
Lys	Leu 370	Lys	Glu	Ile	Leu	Glu 375	Asp	Pro	Gly	Ala	Lys 380		Val	Gly	Gln	
AAT	TTG	AAA	TTC	GAT	TAC	AAG	GTG	TTG	ATG	GTG	AAG	GGT	GTT	GAA	CCT	1200
Asn 385	Leu	Lys	Phe	Asp	Tyr 390	Lys	Val	Leu	Met	Val 395	Lys	Gly	Val	Glu	400	
GTT	CCT	CCT	TAC	TTC	GAC	ACG	ATG	ATA	CCC	GCT	TAC	CTŢ	CTT	GAG	CCG	1248
Val	Pro	Pro	Tyr	Phe 405	Asp	Thr	Met	Ile	Ala 410	Ala	Tyr	Leu	Leu	Glu 415	Pro	
AAC	GAA	AAG	AAG	TTC	AAT	CTG	GAC	GAT	CTC	GCA	TTG	AAA	TTT	CTT	GGA	1296
Asn	Glu	Lys	Lys 420	Phe	Asn	Leu	Asp	Asp 425	Leu	Ala	Leu	Lys	Phe 430	i.eu	Gly	
TAC	AAA	ATG	ACA	TCT	TAC	CAA	GAG	CTC	ATG	TCC	TTC	TCT	TTT	CCG	CTG	1344
Tyr	Lys	Met 435	Thr	Ser	Tyr	Gln	Glu 440	Ļeu	Met	Ser	Phe	Ser 445	Phe	Pro	Leu	
TTT	GGT	TTC	AGT	TTT	GCC	GAT	GTT	CCT	GTA	GAA .	AAA	GCA	GCG	AAC	TAC	1392
	Gly 450		Ser	Phe	Ala	Asp 455	Val	Pro	Val	Glu	Lys 460	Ala	Ala	Asn	Tyr	

1440	ACC CTG AGC	TAC AAG	CTT	AGA	TAC	ACC	ATC	GAG	GCA	GAT	GAA	TGT	TCC
	Thr Leu Ser 480		Leu 475	Arg	Tyr	Thr		Asp 470	Ala	Asp	Glu	Cys	Ser 465
1488	AAG ATA GAA	TTC TAC	GTG	AAC	GAA	CTG	GAT	GCA	GAG	CAC	CTC	AAA	TTA
	Lys Ile Glu 495	Phe Tyr		490	Glu	Leu	Asp		Glu 485	His	Leu	Lys	Leu
1536	GGT GTG TAT	CTG AAC	GAA	ATG	CGG	GCA	CTT	GTG	AAC	GTG	CTT	CCC	ATG
	Gly Val Tyr 510	Leu Asn	Glu		Arg 505	Ala	Leu	Val	Asn	Val 500	Leu	Pro	Met
1584	GGA · AAA AAA	GAG TAC	GAA	TCA	CTC	AAA	AAG	CTG	TTC	GAG	ACA	GAC	GTG
	Gly Lys Lys	Glu Tyr 525	Glu	Ser		Lys 520	Lys	Leu	Phe		Thr 515	Asp	Val
1632	GAG CCG TTC	GCT GGA	ATA	AGG	TAC	ATA	GAA	GAG	GCA	CTG	GAA	GAA	CTC
	Glu Pro Phe	Ala Gly 540	Ile	Arg	Tyr		Glu 535	Glu	Ala	Leu	Glu	Glu 530	Leu
1680	GAA AAA CTC	CTT TTT	ATC	AGG	TCA	GTT	CAG	AAG	CCG	TCA	AAC	ATA	AAC
	Glu Lys Leu 560	Leu Phe	11e 555	Arg	Ser	Val		Lys 550	Pro	Ser	Asn	Ile	Asn 545
1728	TAT TCA ACA	GGA GAC	ACG	AAA	ACG	ACG	AAA	GGT	CGT	CCA	AAA	ATA	GGC
	Tyr Ser Thr 575	Gly Asp	Thr	Lys 570	Thr	Thr	Lys	Gly	Arg 565	Pro	Lys	Ile	Gly
1776	ATC ATT CCT	CAC GAA	GAA	GGT	GCC	CTT	GAA	GAG	CTC	GTC	GAA	ATA	CGC
	Ile Ile Pro 590	His Glu	Glu		Ala 585	Leu	Glu	Glu	Leu	Val 580	Glu	Ile	Arg
1824	ACC TAC ATA	AAA TCA	TTG	AAA	CAG	ATA	AAG	AGA	TAC	GAA	CTT	ATT	CTG
	Thr Tyr Ile	Lys Ser 605	Leu	Lys		11e 600	Lys	Arg	Tyr	Glu	Leu 595	Ile	Leu
1872	ATT CAT GCT	GGA AGG	ACC	AAG	CCA	AAC	GTC	ATG	AAG	CCC	CTT	GCT	GAC
	Ile His Ala	Gly Arg 620	Thr	Lys	Pro	Asn	Val 615	Met	Lys	Pro	Leu	Ala 610	Asp
1920	AGC AGC GAT	CTT AGC	AGA	GGA	ACT	GCC	ACT	GGG	ACG	CAA	AAT	TTC	TCT
	Ser Ser Asp 640	Leu Ser	Arg 635	Gly	Thr	Ala	Thr	G1y 630	Thr	Gln	Asn	Phe	Ser 625

CCC	TAA	CTT	CAG	AAC	CTC	CCG	ACG	AAA	AGT	GAA	GAG	GGA	AAA	GAA A	TC	1968
Pro	Asn	Leu	Gln	Asn 645		Pro	Thr	Lys	5 Ser 650		Glu	ı Gly	Lys	655	Ile	
AGG	AAA	GCG	ATA	GTT	CCT	CAG	GAT	CCA	AAC	TGG	TGG	ATC	GTC	AGT G	CC	2016
Arg	Lys	Ala	Ile 660		Pro	Gln	Asp	Pro 665		Trp	Trp	Ile	Val 670	Ser	Ala	
GAC	TAC	TCC	CAA	ATA	GAA	CTG	AGG	ATC	CTC	GCC	CAT	CTC	AGT	GGT G	AT	2064
Asp	-	Ser 675		Ile	Glu	Leu	Arg 680		Leu	Ala	His	Leu 685		Gly .	Asp	
GAG	AAT	CTT	TTG	AGG	GCA	TTC	GAA	GAG	GGC	ATC	GAC	GTC	CAC	ACT C	TA	2112
Glu	Asn 690	Leu	Leu	Arg	Ala	Phe 695	Glu	Glu	Gly	Ile	Asp 700		His	Thr :	Leu	
ACA	GCT	TCC	AGA	ATA	TTC	AAC	GTG	AAA	ccc	GAA	GAA	GTA	ACC	GAA G	AA	2160
Thr 705	Ala	Ser	Arg	Ile	Phe 710	Asn	Val	Lys	Pro	Glu 715	Glu	Val	Thr	Glu (	Glu 720	
ATG	CGC	CGC	GCT	GGT	AAA	ATG	GTT	AAT	TTT	TCC	ATC	ATA	TAC	GGT G	ΓΑ	2208
Met	Arg	Arg	Ala	Gly 725	Lys	Met	Val	Asn	Phe 730	Ser	Ile	Ile	Tyr	Gly V 735	Val	
ACA	CCT	TAC	GGT	CTG	TCT	GTG	AGG	CTT	GGA	GTA	CCT	GTG .	AAA	GAA GO	CA	2256
Thr	Pro	Tyr	Gly 740	Leu	Ser	Val	Arg	Leu 745	Gly	Val	Pro	Val	Lys 750	Glu A	Ala	
GAA	AAG	ATG	ATC	GTC	AAC	TAC	TTC	GTC	CTC	TAC	CCA	AAG (	GTG (	CGC GA	ΥT	2304
Glu	Lys	Met 755	Ile	Val	Asn	Tyr	Phe 760	Val	Leu	Tyr	Pro	Lys 765	Val	Arg A	Asp	
TAC	TTA	CAG	AGG	GTC	GTA '	TCG	GAA	GCG	AAA (	GAA A	AAA (	GGC :	TAT	GTT AG	SA.	2352
Tyr	Ile 770	Gln	Arg	Val	Val	Ser 775	Glu	Ala	Lys	Glu	Lys 780	Gly	Tyr	Val A	rg	
ACG	CTG	TTT	GGA A	AGA .	AAA A	AGA (	GAC A	ATA	CCA (	CAG (	CTC A	ATG (	GCC (	CCG GA	.c	2400
Thr 785	Leu	Phe	Gly	Arg	Lys 790	Arg	Asp	Ile	Pro	Gln 795	Leu	Met	Ala	Arg A	.sp 00	
AGG .	AAC .	ACA	CAG (	GCT (	GAA (	GGA (	GAA (	CGA .	ATT (	GCC A	ATA A	AAC A	CT C	CC AT	A	2448
Arg	Asn	Thr		Ala 805	Glu	Gly	G1u	Arg	Ile 810	Ala	Ile	Asn	Thr	l·ro I 815	le	

CAG	GGT	ACA	GCA	GCG	GAT	ATA	ATA	AAG	CTG.	GCT	ATG	ATA	GAA A	ATA G	GAC	2496
Gln	Gly	Thr	Ala 820	Ala	Asp	Ile	Ile	Lys 825	Leu	Ala	Met	Ile	Glu 830	Ile	Asp	
AGG	GAA	CTG	AAA	GAA	AGA	AAA	ATG	AGA	TCG	AAG	ATG	ATC	ATA (	CAG G	STC	2544
Arg	Glu	Leu 835	Lys	Glu	Arg	Lys	Met 840		Ser	Lys	Met	11e 845	Ile	Gln	Val	
CAC	GAC	GAA	CTG	GTT	TTT	GAA	GTG	CCC	AAT	GAG	GAA	AAG	GAC C	CC C	TC	2592
His	Asp 850	Glu	Leu	Val	Phe	Glu 855	Val	Pro	Asn	Glu	Glu 860		Asp	Ala	Leu	
GTC	GAG	CTG	GTG	AAA	GAC	AGA	ATG	ACG .	AAT	GTG	GTA	AAG (	CTT 1	CA G	TG	2640
Val 865	Glu	Leu	Val	Lys	Asp 870	Arg	Met	Thr	Asn	Val 875	Val	Lys	Leu		Val 880	
CCG	CTC	GAA	GTG	GAT	GTA	ACC	ATC	GGC .	AAA	ACA	TGG	TCG	TGA			2682
Pro	Leu	Glu	Val	Asp 885	Val	Thr	Ile	Gly	Lys 890	Thr	Trp	Ser			^	

#### (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 893 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ala Arg Leu Phe Leu Phe Asp Gly Thr Ala Leu Ala Tyr Arg Ala 1 5 10 15

Tyr Tyr Ala Leu Asp Arg Ser Leu Ser Thr Ser Thr Gly 11e Pro Thr 20 25 30

Asn Ala Thr Tyr Gly Val Ala Arg Met Leu Val Arg Phe Ile Lys Asp 35 40 45

His Ile Ile Val Gly Lys Asp Tyr Val Ala Val Ala Phe Asp Lys Lys 50 60

Ala Ala Thr Phe Arg His Lys Leu Leu Glu Thr Tyr Lys Ala Gln Arg 65 70 75 80

Pro Lys Thr Pro Asp Leu Leu Ile Gln Gln Leu Pro Tyr Ile Lys Lys 85 90 95

Leu	Val	Glu	Ala 100	Leu	Gly	Met	Lys	Val 105		Glu	Val	Glu	Gly 110	Tyr	Gl
Ala	Asp	Asp 115	Ile	Ile	Ala	Thr	Leu 120	Ala	Val	Lys	Gly	Leu 125	Pro	Leu	Phe
Asp	Glu 130	Ile	Phe	Ile	Val	Thr 135	Gly	Asp	Lys	Asp	Met 140	Leu	Gln	Leu	Va:
Asn 145	Glu	Lys	Ile	Lys	Val 150	Trp	Arg	Ile	Val	Lys 155	Gly	lle	Ser	Asp	Let 160
Glu	Leu	Tyr	Asp	Ala 165	Gln	Lys	Val	Lys	Glu 170	Lys	Tyr	Gly	Val	Glu 175	Pro
			180		Leu			185					190		
		195			Gly		200					205			
	210				Leu	215					220				
225		_			Lys 230					235					240
				245	Ala				250					255	
	_		260		Arg			265					270		
		275			Leu		280					285			
	290				Glu	295					300				
305					Leu 310					315					320
				325	Thr				330					335	
			340		Ser			345					350		
		355			Ala		360					365			
-	370				Leu	375					380				
Asn 385	Leu	Lys	Phe	Asp	Tyr 390	Lys	val	Leu	met	Val 395	ьys	сту	val	"IU	400

Va.	L Pro	) Pr	о Туі	Phe 405	Asp	Thi	r Met	: Ile	≥ Ala 410		а Туг	r Lei	u Lei	41:	
Ası	ı Glı	ı Ly:	420	Phe	e Asr	Let	ı Asp	425	Let	ı Ala	i Let	ı Lys	s Phe 430		u Gly
Tyr	Lys	435	t Thr	Ser	Tyr	Glī	440	Let	1 Met	: Sei	Phe	Ser 445		e Pro	Let
Phe	450	Phe	e Ser	Phe	Ala	455	Val	Pro	Va]	. Glu	Lys 460		Ala	Asr	туг
Ser 465	Cys	Glu	ı Asp	Ala	470	Ile	Thr	Tyr	Arg	475		Lys	Thr	Leu	Ser 480
			His	485					490	•				495	•
			Val 500					505					510		
		212					520					525			
	230		Leu			535					540				
343	•		Ser		550					555					560
			Pro	565					570					575	
			Val 580					585					590		
		595	Glu				600					605			
	610		Pro			615					620				
Ser 625	Phe	Asn	Gln	Thr	Gly 630	Thr	Ala	Thr	Gly	Arg 635	Leu	Ser	Ser	Ser	Asp 640
Pro	Asn	Leu	Gln	Asn 645	Leu	Pro	Thr	Lys	Ser 650	Glu	Glu	Gly	Lys	G1u 655	Ile
Arg	Lys	Ala	Ile 660	Val	Pro	Gln	Asp	Pro 665	Asn	Trp	Trp	Ile	Val 670	Ser	Ala
Asp	Tyr	Ser 675	Gln	Ile	Glu	Leu .	Arg 680	Ile	Leu	Ala		Leu 685	Ser	Gly	Asp

Glu Asn Leu Leu Arg Ala Phe Glu Glu Gly Ile Asp Val His Thr Leu 690 695 700

Thr Ala Ser Arg Ile Phe Asn Val Lys Pro Glu Glu Val Thr Glu Glu 705 710 715 720

Met Arg Arg Ala Gly Lys Met Val Asn Phe Ser Ile Ile Tyr Gly Val 725 730 735

Thr Pro Tyr Gly Leu Ser Val Arg Leu Gly Val Pro Val Lys Glu Ala 740 745 750

Glu Lys Met Ile Val Asn Tyr Phe Val Leu Tyr Pro Lys Val Arg Asp
755 760 765

Tyr Ile Gln Arg Val Val Ser Glu Ala Lys Glu Lys Gly Tyr Val Arg 770 780

Thr Leu Phe Gly Arg Lys Arg Asp Ile Pro Gln Leu Met Ala Arg Asp 785 790 795 800

Arg Asn Thr Gln Ala Glu Gly Glu Arg Ile Ala Ile Asn Thr Pro Ile 805 810 815

Gln Gly Thr Ala Ala Asp Ile Ile Lys Leu Ala Met Ile Glu Ile Asp 820 825 830

Arg Glu Leu Lys Glu Arg Lys Met Arg Ser Lys Met Ile Ile Gln Val 835 840 845

His Asp Glu Leu Val Phe Glu Val Pro Asn Glu Glu Lys Asp Æla Leu 850 860

Val Glu Leu Val Lys Asp Arg Met Thr Asn Val Val Lys Leu Ser Val 865 870 875 880

Pro Leu Glu Val Asp Val Thr Ile Gly Lys Thr Trp Ser 885 890

#### (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2493 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Thermus species sps17

## (ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..2490

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG	CTG	CCC	CTC	TTT	GAG	ccc	AAG	GGC	CGG	GTC	CTC	CTG	GTG	GAC	GGC	48
Met 1	Leu	Pro	Leu	Phe 5		Pro	Lys	Gly	Arg		Leu	ı Leı	ı Va	1 As <sub>1</sub>	p Gly 5	
CAC	CAC	CTG	GCC	TAC	CGC	ACC	TTT	TTC	GCC	CTC	AAG	GGC	CTC	ACC	ACC	96
His	His	Leu	Ala 20	Tyr	Arg	Thr	Phe	Phe 25	Ala	Leu	Lys	Gly	/ Lei		r Thr	
AGC	CGG	GGC	GAG	CCC	GTG	CAG	GCG	GTT	TAT	GGC	TTC	GCC	AAA	AGC	CTC	144
Ser	Arg	Gly 35	Glu	Pro	Val	Gln	Ala 40		Tyr	Gly	Phe	Ala 45		s Sei	Leu	
CTC	AAG	GCC	CTG	AAG	GAG	GAT	GGG	GAG	GTG	GCC	ATC	GTG	GTC	TTT	GAC	192
Leu	Lys 50	Ala	Leu	Lys	Glu	Asp 55	Gly	Glu	Val	Ala	Ile 60	Val	. Val	l Phe	e Asp	
GCC	AAG	GCC	CCC	TCC	TTC	CGC	CAC	GAG	GCC	TAC	GAG	GCC	TAC	AAG	GCG	240
Ala 65	Lys	Ala	Pro	Ser	Phe 70	Arg	His	G1u	Ala	Tyr 75	Glu	Ala	Туг	Lys	80	
GGC	CGG	GCC	CCC	ACC	CCG	GAG	GAC	TTT	ccc	CGG	CAG	CTC	GCC	CTC	ATC	288
Gly	Arg	Ala	Pro	Thr 85	Pro	Glu	Asp	Phe	Pro 90	Arg	Gln	Leu	Ala	Leu 95	lle	
AAG	GAG	CTG	GTG	GAC	CTT	TTG	GGC	CTC	GTG	CGC	CTT	GAG	GTC	CCG	GGC	336
Lys	Glu	Leu	Val 100	Asp	Leu	Leu	Gly	Leu 105	Val	Arg	Leu	Glu	Val		Gly	
TTT	GAG	GCG	GAC	GAT	GTC	CTC	GCC	ACC	CTG	GCC	AAG	AAG	GCA	GAA	AGG	384
Phe	Glu	Ala 115	Asp	Asp	Val	Leu	Ala 120	Thr	Leu	Ala	Lys	Lys 125		Clu	Arg	
GAG	GGG	TAC	GAG	GTG	CGC	ATC	CTG	AGC	GCG	GAC	CGC	GAC	CTC	TAC	CAG	432
Glu	Gly 130	Tyr	Glu	Val	Arg	Ile 135	Leu	Ser	Ala	Asp	Arg 140	Asp	Leu	l'yr	Gln	
CTC	CTT	TCC	GAC	CGG .	ATC	CAC	СТС	CTC	CAC	CCC	GAG	GGG	GAG	GTC	CTG	480
Leu 145	Leu	Ser	Asp	Arg	Ile 150	His	Leu	Leu	His	Pro 155	Glu	Gly	Glu	Val	Leu 160	

ACC	CCC	GGG	TGG	CTC	CAG	GAG	CGC	TAC	GGC	CTC	TCC	CCG	GAG	AGG	TGG	528
Thr	Pro	Gly	Trp	Leu 165	Gln	Glu	Arg	Tyr	Gly 170		Ser	Pro	Glu	175	Trp	
GTG	GAG	TAC	CGG	GCC	CTG	GTG	GGG	GAC	CCT	TCG	GAC	AAC	CTC	CCC	GGG	576
Val	Glu	Tyr	Arg 180	Ala	Leu	Val	Gly	Asp 185	Pro	Ser	Asp	Asn	Leu 190	Pro	Gly	
GTG	CCC	GGC	ATC	GGG	GAG	AAG	ACC	GCC	CTG	AAG	CTC	CTG	AAG	GAG	TGG	624
Val	Pro	Gly 195	Ile	Gly	Glu	Lys	Thr 200	Ala	Leu	Lys	Leu	Leu 205		Glu	Trp	
GGT	AGC	CTG	GAA	GCG	ATT	CTA	AAG	AAC	CTG	GAC	CAG	GTG	AAG	CCG	GAA	672
Gly	Ser 210	Leu	Glu	Ala	Ile	Leu 215	Lys	Asn	Leu	Asp	Gln 220		Lys	Pro	Glu	
AGG	GTG	CGG	GAG	GCC	ATC	CGG	TAA	AAC	CTG	GAT	AAG	CTC	CAG	ATG	TCC	720
Arg 225	Val	Arg	Glu	Ala	Ile 230	Arg	Asn	Asn	Leu	Asp 235		Leu	Gln	Met	Ser 240	
CTG	GAG	CTT	TCC	CGC	CTC	CGC	ACC	GAC	CTC	CCC	CTG	GAG	GTG	GAC	TTC	768
Leu	Glu	Leu	Ser	Arg 245	Leu	Arg	Thr	Asp	Leu 250	Pro	Leu	Glu	Val	Asp 255	Phe	
GCC	AAG	AGG	CGG	GAG	CCC	GAC	TGG	GAG	GGG	CTT	AAG	GCC	TTT	TTG	GAG	816
Ala	Lys	Arg	Arg 260	Glu	Pro	Asp	Trp	Glu 265	Gly	Leu	Lys	Ala	Phe 270	Leu	Glu	
CGG	CTT	GAG	TTC	GGA	AGC	CTC	CTC	CAC	GAG	TTC	GGC	CTT	CTG	GAG	GCC	864
Arg	Leu	Glu 275	Phe	Gly	Ser	Leu	Leu 280	His	Glu	Phe	Gly	Leu 285		Clu	Ala	
CCC	AAG	GAG	GCG	GAG	GAG	GCC	ccc	TGG	CCC	CCG	CCT	GGA	GGG	GCC	TTT	912
Pro	Lys 290	Glu	Ala	Glu	Glu	Ala 295	Pro	Trp	Pro	Pro	Pro 300	Gly	Gly	Ala	Phe	
TTG	GGC	TTC	CTC	CTC	TCC	CGC	ссс	GAG	CCC .	ATG	TGG	GCG	GAG	CTT	TTG	960
Leu 305	Gly	Phe	Leu	Leu	Ser 310	Arg	Pro	Glu	Pro	Met 315	Trp	Ala	Glu	Leu	Leu 320	
GCC	CTG	GCG	GGG	GCC	AAG	GAG	GGG	CGG	GTC	CAT	CGG	GCG	GAA (	GAC (	ccc	1008
Ala	Leu	Ala	Gly	Ala	Lys	Glu	Gly	Arg	Val	His	Arg	Ala	Glu	<i>i</i> .sp 335	Pro	

GTG	GGG	GCC	CTA	AAG	GAC	CTG	AAG	GAG	ATC	CGG	GGC	CTC	CTC	GCC	AAG	1056
Val	Gly	Ala	Leu 340	Lys	Asp	Leu	Lys	G1u 345		Arg	.Gly	Leu	Leu 350		Lys	
GAC	CTC	TCG	GTC	CTG	GCC	CTG	AGG	GAG	GGC	CGG	GAG	ATC	CCG	CCG	GGG	1104
Asp	Leu	Ser 355	Val	Leu	Ala	Leu	Arg 360		Gly	Arg	Glu	11e 365		Pro	Gly	
GAC	GAC	CCC	ATG	CTC	CTC	GCC	TAC	CTC	CTG	GAC	CCG	GGG	AAC	ACC	AAC	1152
Asp	Asp 370	Pro	Met	Leu	Leu	Ala 375	Tyr	Leu	Leu	Asp	Pro 380	_	' Asn	Thr	Asn	
ccc	GAG	GGG	GTG	GCC	CGG	CGG	TAC	GGG	GGG	GAG	TGG	AAG	GAG	GAC	GCC	1200
Pro 385	Glu	Gly	Val	Ala	Arg 390	Arg	Tyr	Gly	Gly	Glu 395	-	Lys	Glu	Asp	Ala 400	
GCC	GCC	CGG	GCC	CTC	CTT	ŢCG	GAA	AGG	CTC	TGG	CAG	GCC	CTT	TAC	ccc	1248
Ala	Ala	Arg	Ala	Leu 405	Leu	Ser	Glu	Arg	Leu 410		Gln	Ala	Leu	Tyr 415		
CGG	GTG	GCG	GAG	GAG	GAA	AGG	CTC	CTT	TGG	CTC	TAC	CGG	GAG	GTG	GAG	1296
Arg	Val	Ala	Glu 420	Glu	Glu	Arg	Leu	Leu 425	Trp	Leu	Tyr	Arg	Glu 430		Glu	
CGG	ccc	CTC	GCC	CAG	GTC	CTC	GCC	CAC	ATG	GAG	GCC	ACG	GGG	GTG	CGG	1344
Arg	Pro	Leu 435	Ala	Gln	Val	Leu	Ala 440	His	Met	Glu	Ala	Thr 445	Gly	Val	Arg	
CTG	GAT	GTG	CCC	TAC	CTG	GAG	GCC	CTT	TCC	CAG	GAG	GTG	GCC	T:T	GAG	1392
Leu	Asp 450	Val	Pro	Tyr	Leu	Glu 455	Ala	Leu	Ser	Gln	Giu 450		Ala	he	Glu	
CTG	GAG	CGC	CTC	GAG	GCC	GAG	GTC	CAC	CGC	CTG	GCG	GGC	CAC	CCC	TTC	1440
Leu 465	Glu	Arg	Leu	Glu	Ala 470	Glu	Val	His	Arg	Leu 475		Gly	His	Pro	Phe 480	
AAC	CTG	AAC	TCT	AGG	GAC	CAG	CTG	GAG	CGG	GTC	CTC	TTT	GAC	GAG	CTC	1488
Asn	Leu	Asn	Ser	Arg 485	Asp	Gln	Leu	Glu	Arg 490	Val	Leu	Phe	Asp	Glu 495		
GGC	CTA	CCC	ccc	ATC	GGC	AAG	ACG	GAG	AAG	ACG	GGC	AAG	CGC	TCC	ACC	1536
Gly	Leu	Pro	Pro	Ile	Gly	Lys	Thr	Glu 505	-	Thr	Gly	Lys	Arg		Thr	

AGC	GCC	GCC	GTC	CTG	GAG	CTC	TTA	AGG	GAG	GCC	CAC	CCC	ATC	GTG	GGG	1584
Ser	Ala	Ala 515		Leu	Glu	Leu	Leu 520		Glu	ı Ala	a His	525		e Va	1 Gly	7
CGG	ATC	CTG	GAG	TAC	CGG	GAG	CTC	ATG	AAG	CTC	AAG	AGC	ACC	TAC	ATA	1632
Arg	Ile 530		Glu	Tyr	Arg	Glu 535		Met	Lys	Le	u Lys 540		c Th	r Ty	r Ile	<b>:</b>
GAC	CCC	CTC	CCC	AGG	CTG	GTC	CAC	CCC	AAA	ACC	GGC	CGG	CTC	CAC	AGC	1680
Asp 545		Leu	Pro	Arg	Leu 550		His	Pro	Lys	Th: 555		7 Arg	g Le	u Hi	s Thr 560	•
CGC	TTC	AAC	CAG	ACG	GCC	ACC	GCC	ACG	GGC	CGC	CTC	TCC	AGC	TCC	GAC	1728
Arg	Phe	Asn	Gln	Thr 565	Ala	Thr	Ala	Thr	Gly 570		g Lev	ı Ser	Se	r Se 57	r Asp 5	
CCC	AAC	CTG	CAG	AAC	ATC	CCC	GTG	CGC	ACC	CCC	TTA	GGC	CAG	CGC	ATC	1776
Pro	Asn	Leu	Gln 580	Asn	Ile	Pro	Val	Arg 585	Thr	Pro	Leu	Gly	Gl <sub>1</sub> 590		g Ile	
CGC	AAG	GCC	TTC	ATT	GCC	GAG	GAG	GGC	CAT	CTC	CTG	GTG	GCC	CTG	GAC	1824
Arg	Lys	Ala 595	Phe	Ile	Ala	Glu	Glu 600		His	Lev	ı Leu	Val 605		a Lei	u Asp	
TAT	AGC	CAG	ĀTC	GAG	CTC	CGG	GTC	CTC	GCC	CAC	CTC	TCG	GGG	GAC	GAG	1872
Tyr	Ser 610	Gln	Ile	Glu	Leu	Arg 615	Val	Leu	Ala	His	620		Gly	7 Asj	Glu	
AAC	CTC	ATC	CGG	GTC	TTC	CGG	GAA	GGG	AAG	GAC	ATC	CAC	ACC	GAG	ACC	1920
Asn 625	Leu	Ile	Arg	Val	Phe 630	Arg	Glu	Gly	Lys	Asp 635		His	Thi	Glu	1 Thr 640	
GCC	GCC	TGG	ATG	TTC	GGC	GTG	ccc	ССС	GAG	GGG	GTG	GAC	GGG	GCC	ATG	1968
Ala	Ala	Trp	Met	Phe 645	Gly	Val	Pro		Glu 650		Val	Asp	Gly	Ala 655	Met	
CGC	CGG	GCG	GCC	AAG	ACG	GTG	AAC	TTC	GGG	GTG	CTC	TAC	GGG	ATG	TCC	2016
Arg	Arg <sub>.</sub>	Ala	Ala 660	Lys	Thr	Val	Asn	Phe 665	Gly	Val	Leu	Tyr	Gly 670		Ser	
GCC	CAC	CGC	CTC	TCC	CAG	GAG	CTC	TCC .	ATC	ccc	TAC	GAG	GAG	GCG	GCG	2064
Ala	His	Arg	Leu	Ser	Gln	Glu	Leu 680	Ser	Ile	Pro	Tyr	Glu 685	Glu	Ala	Ala	

GCC	TTC	ATC	GAG	CGC	TAC	TTC	CAG	AGC	TTC	ccc	AAG	GTG	CGG	GCC	TGG	2112
Ala	Phe 690	Ile	: Glu	ı Arg	ту1	Phe 695		Ser	Phe	Pro	700		. Arg	g Ala	Trp	
ATC	GCC	AAA	ACC	TTG	GAG	GAG	GGG	CGG	AAG	AAG	GGG	TAC	GTG	GAG .	ACC	2160
Ile 705	Ala	Lys	Thr	Leu	Glu 710	Glu	Gly	Arg	Lys	Lys 715		Tyr	Val	. Glu	Thr 720	
CTC	TTC	GGC	CGC	CGC	CGC	TAC	GTG	CCC	GAC	CTC	AAC	GCC	CGG	GTG .	AAG	2208
Leu	Phe	G1y	Arg	Arg 725		Tyr	Val	Pro	Asp 730		Asn	Ala	Arg	735	Lys	
AGC	GTG	CGG	GAG	GCG	GCG	GAG	CGC	ATG	GCC	TTC	AAC	ATG	CCC	GTG (	CAG	2256
Ser	Val	Arg	Glu 740	Ala	Ala	Glu	Arg	Met 745	Ala	Phe	Asn	Met	Pro 750	Val	Gln	
GGC	ACC	GCC	GCG	GAC	CTC	ATG	AAG	CTG	GCC	ATG	GTG	AAG	CTC	TTC (	ccc	2304
Gly	Thr	Ala 755	Ala	Asp	Leu	Met	Lys 760	Leu	Ala	Met	Val	Lys 765	Leu	Phe	Pro	
AGG	CTC	AGG	CCC	TTG	GGC	GTT	CGC	ATC	CTC	CTC	CAG	GTG	CAC	GAC (	GAG	2352
Arg	Leu 770	Arg	Pro	Leu	Gly	Val 775	Arg	Ile	Leu	Leu	Gln 780	Val	His	Asp	Glu	
CTG	GTC	TTG	ĢAG	GCC	CCA	AAG	GCG	CGG	GCG	GAG	GAG	GCC (	GCC	CAG 1	TG	2400
Leu 785	Val	Leu	Glu	Ala	Pro 790	Lys	Ala	Arg	Ala	Glu 795	G1u	Ala	Ala	Gln	Leu 800	
GCC	AAG	GAG	ACC	ATG	GAA	GGG	GTT	TAC	CCC	CTC	TCC	GTC (	CCC	CTG G	AG	2448
Ala	Lys	Glu	Thr	Met 805	Glu	Gly	Val	Tyr	Pro 810	Leu	Ser	Val	Pro	Leu 815	Glu	
GTG	GAG	GTG	GGG	ATG	GGG	GAG	GAC	TGG	CTT	TCC	GCC .	AAG (	GCC			2490
Val	Glu	Val	Gly 820	Met	Gly	Glu	Asp	Trp 825	Leu	Ser	Ala	Lys	Ala 830			
TAG																2493

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 830 amino acids
  (B) TYPE: amino acid
  (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: protein

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val Asp Gly

1 5 10 15

His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu Thr Thr 20 25 30

Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys Ser Leu 35 40 45

Leu Lys Ala Leu Lys Glu Asp Gly Glu Val Ala Ile Val Val Phe Asp 50 60

Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu Ala Tyr Lys Ala 65 70 75 80

Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu Ala Leu Ile 85 90 95

Lys Glu Leu Val Asp Leu Leu Gly Leu Val Arg Leu Glu Val Pro Gly
100 105 110

Phe Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Lys Ala Glu Arg 115 120 125

Glu Gly Tyr Glu Val Arg Ile Leu Ser Ala Asp Arg Asp Leu Tyr Gln 130 135 140

Leu Leu Ser Asp Arg Ile His Leu Leu His Pro Glu Gly Glu Val Leu 145 150 155 160

Thr Pro Gly Trp Leu Gln Glu Arg Tyr Gly Leu Ser Pro Glu Arg Trp
165 170 175

Val Glu Tyr Arg Ala Leu Val Gly Asp Pro Ser Asp Asn Leu Pro Gly
180 185 190

Val Pro Gly Ile Gly Glu Lys Thr Ala Leu Lys Leu Leu Lys Glu Trp 195 200 205

Gly Ser Leu Glu Ala Ile Leu Lys Asn Leu Asp Gln Val Lys Pro Glu 210 215 220

Arg Val Arg Glu Ala Ile Arg Asn Asn Leu Asp Lys Leu Gln Met Ser 225 230 235 240

Leu Glu Leu Ser Arg Leu Arg Thr Asp Leu Pro Leu Glu Val Asp Phe 245 250 255

Ala Lys Arg Glu Pro Asp Trp Glu Gly Leu Lys Ala Phe Leu Glu 260 265 270

Arg	Leu	Glu 275	Phe	Gly	Ser	Leu	Leu 280	His	Glu	Phe	Gly	Leu 285	Leu	Glu	Ala
Pro	Lys 290	Glu	Ala	Glu	Glu	Ala 295	Pro	Trp	Pro	Pro	Pro 300	Gly	Gly	Ala	Phe
Leu 305	Gly	Phe	Leu	Leu	Ser 310	Arg	Pro	Glu	Pro	Met 315	Trp	Ala	Glu	Leu	Leu 320
Ala	Leu	Ala	Gly	Ala 325	Lys	Glu	Gly	Arg	Val 330	His	Arg	Ala	Glu	Asp 335	Pro
Val	Gly	Ala	Leu 340	Lys	Asp	Leu	Lys	Glu 345	Ile	Arg	Gly	Leu	Leu 350	Ala	Lys
Asp	Leu	Ser 355	Val	Leu	Ala	Leu	Arg 360	Glu	Gly	Arg	Glu	Ile 365	Pro	Pro	Gly
Asp	Asp 370	Pro	Met	Leu	Leu	Ala 375	Tyr	Leu	Leu	Asp	Pro 380	Gly	Asn	Thr	Asn
Pro 385	Glu	Gly	Val	Ala	Arg 390	Arg	Tyr	Gly	G1y	Glu 395	Trp	Lys	Glu	Asp	Ala 400
Ala	Ala	Arg	Ala	Leu 405	Leu	Ser	Glu	Arg	Leu 410	Trp	Gln	Ala	Leu	Tyr 415	Pro
Arg	Val	Ala	Glu 420	Glu	Glu	Arg	Leu	Leu 425	Trp	Leu	Tyr	Arg	Glu 430	Val	Glu
Arg	Pro	Leu 435	Ala	Gln	Val	Leu	Ala 440	His	Met	Glu	Ala	Thr 445	Gly	Val	Arg
Leu	Asp 450	Val	Pro	Tyr	Leu	Glu 455	Ala	Leu	Ser	Gln	Glu 460	Val	Ala	Phe	Glu
Leu 465	Glu	Arg	Leu	Glu	Ala 470	Glu	Val	His	Arg	Leu 475	Ala	G1y	His	Pro	Phe 480
Asn	Leu	Asn	Ser	Arg 485	Asp	Gln	Leu	Glu	Arg 490	Val	Leu	Phe	Asp	Glu 495	Leu
Gly	Leu	Pro	Pro 500	Ile	Gly	Lys		Glu 505	Lys	Thr	Gly	Lys	Arg 510	Ser	Thr
Ser	Ala	Ala 515	Val	Leu	Glu	Leu	Leu 520	Arg	Glu	Ala	His	Pro 525	Ile	Val	Gly
Arg	Ile 530	Leu	Glu	Tyr	Arg	Glu 535	Leu	Met	Lys	Leu	Lys 540	Ser	Thr	Tyr	Ile
Asp 545	Pro	Leu	Pro	Arg	Leu 550	Val	His	Pro	Lys	Thr 555	Gly	Arg	Leu	His	Thr 560
Arg	Phe	Asn	Gln	Thr 565	Ala	Thr	Ala	Thr	Gly 570	Arg	Leu	Ser	Ser	Ser 575	Asp

 Pro
 Asn
 Leu
 Gln
 Asn
 Ile
 Pro
 Val
 Arg
 Thr
 Pro
 Leu
 Gly
 Gln
 Arg
 Ile
 Seps
 Thr
 Fro
 Leu
 Leu
 Val
 Leu
 Asp
 Leu
 Asp
 Glu
 Asp
 Glu</t

Arg Leu Arg Pro Leu Gly Val Arg Ile Leu Leu Gln Val His Asp Glu
770 780

Leu Val Leu Glu Ala Pro Lys Ala Arg Ala Glu Glu Ala Ala Gln Leu 785 790 795 800

Ala Lys Glu Thr Met Glu Gly Val Tyr Pro Leu Ser Val Pro Leu Glu 805 810 815

Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala Lys Ala 820 825 830

#### (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2505 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear		
(ii) MOLECULE TYPE: DNA (genomic)		
(iii) HYPOTHETICAL: NO		
(iv) ANTI-SENSE: NO		
<pre>(vi) ORIGINAL SOURCE:    (A) ORGANISM: Thermus species ZO5</pre>		
(ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 12502		
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:		
ATG AAG GCG ATG CTT CCG CTC TTT GAA CCC AAA GGC CGG GTT CTC CTG	48	
Met Lys Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu 1 5 10 15		
GTG GAC GGC CAC CTG GCC TAC CGC ACC TTC TTC GCC CTA AAG GGC	96	
Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly 20 25 30		
CTC ACC ACG AGC CGG GGC GAA CCG GTG CAG GCG GTT TAC GGC TTC GCC	144	
Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala 35 40 45		
AAG AGC CTC CTC AAG GCC CTG AAG GAG GAC GGG TAC AAG GCC GTC TTC	192	
Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe 50 55 60		
GTG GTC TTT GAC GCC AAG GCC CCT TCC TTC CGC CAC GAG GCC TAC GAG	240	
Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu 65 70 75 80		
GCC TAC AAG GCA GGC CGC GCC CCG ACC CCC GAG GAC TTC CCC CGG CAG	288	
Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln 85 90 95		
CTC GCC CTC ATC AAG GAG CTG GTG GAC CTC CTG GGG TTT ACT CGC CTC	336	:

GAG	GTI	CCG	GGC	TTI	GAG	GCG	GAC	GAC	GTC	CTC	GCC	CACC	CTG	GCC	AAG	384
Glu	ı Val	115		Phe	e Glu	ı Ala	120	_	v Va	l Le	u Al	a Th		u Al	a Lys	
AAG	GCG	GAA	AGG	GAG	GGG	TAC	GAG	GTG	CGC	ATC	CTO	ACC	GCC	GAC	CGG	432
Lys	Ala 130		Arg	Glu	ı Gly	Tyr 135		ı Va]	l Ar	g Ile	e Le 14	_	r Ala	a As	p Arg	
GAC	CTT	TAC	CAG	CTC	GTC	TCC	GAC	CGC	GTC	GCC	GTC	CTC	CAC	ccc	GAG	480
Asp 145		Tyr	Gln	Leu	Val 150		Asp	Arg	, Val	1 Ala 155	_	l Lev	ı His	s Pro	0 Glu 160	
GGC	CAC	CTC	ATC	ACC	CCG	GAG	TGG	CTT	TGG	GAG	AAG	TAC	GGC	CTT	AAG	528
Gly	His	Leu	Ile	Thr 165		Glu	Trp	Leu	170		ı Ly:	s <b>Т</b> уг	Gly	7 Let 175	ı Lys	
CCG	GAG	CAG	TGG	GTG	GAC	TTC	CGC	GCC	CTC	GTG	GGG	GAC	CCC	TCC	GAC	576
Pro	Glu	Gln	Trp 180	Val	Asp	Phe	Arg	Ala 185		Val	Gly	y Asp	Pro 190		Asp	
AAC	CTC	CCC	GGG	GTC	AAG	GGC	ATC	GGG	GAG	AAG	ACC	GCC	CTC	AAG	CTC	624
Asn	Leu	Pro 195	Gly	Val	Lys	Gly	lle 200	-	Glu	Lys	Thi	205		Lys	Leu	
CTC	AAG	GAG	TGG	GGA	AGC	CTG	GAA	AAT	ATC	CTC	AAG	AAC	CTG	GAC	CGG	672
Leu	Lys 210	Glu	Trp	Gly	Ser	Leu 215	Glu	Asn	Ile	Leu	Lys 220		Leu	Asp	Arg	
GTG	AAG	CCG	GAA	AGC	GTC	CGG	GAA	AGG	ATC	AAG	GCC	CAC	CTG	GAA	GAC	720
Val 225	Lys	Pro	Glu	Ser	Val 230	Arg	Glu	Arg	Ile	Lys 235		His	Leu	Clu	Asp 240	
CTT	AAG	CTC	TCC	TTG	GAG	CTT '	TCC	CGG	GTG	CGC	TCG	GAC	CTC	CCC	CTG	768
Leu	Lys	Leu	Ser	Leu 245	Glu	Leu	Ser	Arg	Val 250	Arg	Ser	Asp	Leu	Pro 255	Leu	
GAG	GTG	GAC	TTC	GCC	CGG .	AGG (	CGG	GAG	CCT	GAC	CGG	GAA (	GGG (	CTT (	CGG	816
Glu	Val	Asp	Phe 260	Ala	Arg	Arg	Arg	Glu 265	Pro	Asp	Arg	Glu	Gly 270	ï.eu	Arg	
GCC	TTT	TTG	GAG (	CGC	TTG (	GAG 1	TC (	GGC A	AGC 4	CTC	CTC	CAC (	GAG :	TTC (	GGC	864
Ala		Leu 275	Glu .	Arg	Leu		Phe 280	Gly	Ser	Leu	Leu	His 285	Glu	Phe	Gly	

CTC	CTC	GAG	GCC	CCC	GCC	CCC	CTG	GAG	GAG	GCC	CCC	TGG	CCC	CCG	CCG	912
Leu	Leu 290	Glu	Ala	Pro	Ala	Pro 295		Glu	Glu	Ala	Pro 300		Pro	Pro	Pro	
GAA	GGG	GCC	TTC	GTG	GGC	TTC	GTC	CTC	TCC	CGC	CCC	GAG	CCC	ATG	TGG	960
Glu 305	Gly	Ala	Phe	Val	Gly 310		Val	Leu	Ser	Arg 315		Glu	Pro	Met	Trp 320	
GCG	GAG	CTT	AAA	GCC	CTG	GCC	GCC	TGC	AAG	GAG	GGC	CGG	GTG	CAC	CGG	1008
Ala	Glu	Leu	Lys	Ala 325	Leu	Ala	Ala	Cys	Lys 330	Glu	Gly	Arg	, Val	His 335	Arg	
GCA	AAG	GAC	CCC	TTG	GCG	GGG	CTA	AAG	GAC	CTC	AAG	GAG	GTC	CGA	GGC	1056
Ala	Lys	Asp	Pro 340	Leu	Ala	Gly	Leu	Lys 345	Asp	Leu	Lys	Glu	Val 350	_	Gly	
CTC	CTC	GCC	AAG	GAC	CTC	GCC	GTT	TTG	GCC	CTT	CGC	GAG	GGG	CTG	GAC	1104
Leu	Leu	Ala 355	Lys	Asp	Leu	Ala	Val 360	Leu	Ala	Leu	Arg	Glu 365	-	Leu	Asp	
CTC	GCG	CCT	TCG	GAC	GAC	ссс	ATG	CTC	CTC	GCC	TAC	CTC	CTG	GAC	ccc	1152
Leu	Ala 370	Pro	Ser	Asp	Asp	Pro 375	Met	Leu	Leu	Ala	Tyr 380	Leu	Leu	Asp	Pro	
ŢCC	AAC	ACC	ACC	CCC	GAG	GGG	GTG	GCC	CGG	CGC	TAC	GGG	GGG	GAG	TGG	1200
Ser 385	Asn	Thr	Thr	Pro	Glu 390	Gly	Val	Ala	Arg	Arg 395	Tyr	Gly	Gly	Glu	Trp 400	
ACG	GAG	GAC	GCC	GCC	CAC	CGG	GCC	CTC	CTC	GCC	GAG	CGG	CTC	CAG	CAA	1248
Thr	Glu	Asp	Ala	Ala 405	His	Arg	Ala	Leu	Leu 410	Ala	Glu	Arg	Leu	Gln 415	Gln	
AAC	CTC	TTG	GAA	CGC	CTC	AAG	GGA	GAG	GAA .	AAG	CTC	CTT	TGG	CTC :	<b>TAC</b>	1296
Asn	Leu		Glu 420		Leu	Lys		Glu 425		Lys	Leu	Leu	Trp 430	Leu	Tyr	
CAA	GAG	GTG	GAA	AAG	CCC	CTC	TCC	CGG	GTC	CTG	GCC	CAC	ATG	GAG (	GCC	1344
Gln	Glu	Val 435	Glu	Lys	Pro	Leu	Ser 440	Arg	Val	Leu	Ala	His 445	Met	Glu	Ala	
ACC	GGG	GTA	AGG	CTG	GAC	GTG	GCC	TAT	CTA .	AAG	GCC	CTT	TCC	CTG (	GAG	1392
Thr	Gly 450	Val	Arg	Leu	Asp	Val 455	Ala	Tyr	Leu	Lys	Ala 460	Leu	Ser	Leu	Glu	

CTT	GCG	GAG	GAG	ATT	CGC	CGC	CTC	GAG	GAG	GAG	GTC	TTC	CGC	CTG	GCG	1440
Leu 465		Glu	Glu	Ile	Arg 470		Leu	Glu	Glu	Glu 475		Phe	e Arg	g Leu	Ala 480	
GGC	CAC	CCC	TTC	AAC	CTG	AAC	TCC	CGT	GAC	CAG	CTA	GAG	CGG	GTG	CTC	1488
Gly	His	Pro	Phe	Asn 485		Asn	Ser	Arg	Asp 490		Leu	Glu	Arg	Val 495	Leu	
TTT	GAC	GAG	CTT	AGG	CTT	CCC	GCC	CTG	GGC	AAG	ACG	CAA	AAG	ACG	GGG	1536
Phe	Asp	Glu	Leu 500	Arg	Leu	Pro	Ala	Leu 505		Lys	Thr	Gln	Lys 510	Thr	Gly	
AAG	CGC	TCC	ACC	AGC	GCC	GCG	GTG	CTG	GAG	GCC	CTC	AGG	GAG	GCC	CAC	1584
Lys	Arg	Ser 515	Thr	Ser	Ala	Ala	Val 520		Glu	Ala	Leu	Arg 525	Glu	Ala	His	
CCC	ATC	GTG	GAG	AAG	ATC	CTC	CAG	CAC	CGG	GAG	CTC	ACC	AAG	CTC A	AAG	1632
Pro	11e 530	Val	Glu	Lys	Ile	Leu 535	Gln	His	Arg	Glu	Leu 540	Thr	Lys	Leu	Lys	
AAC	ACC	TAC	GTG	GAC	ccc	CTC	CCG	GGC	CTC	GTC	CAC	CCG	AGG	ACG (	GGC	1680
Asn 545	Thr	Tyr	Val	Asp	Pro 550	Leu	Pro	Gly	Leu	Val 555	His	Pro	Arg	Thr	Gly 560	
CGC	CTC	CAC	ACC	CGC	TTC	AAC	CAG	ACA	GCC	ACG	GCC	ACG	GGA .	AGG (	CTC	1728
Arg	Leu	His	Thr	Arg 565	Phe	Asn	Gln	Thr	Ala 570	Thr	Ala	Thr	Gly	Arg 575	Leu	
TCT	AGC	TCC	GAC	CCC	AAC	CTG	CAG	AAC	ATC	CCC A	ATC	CGC	ACC	CCC 1	TTG	1776
Ser	Ser	Ser	Asp 580	Pro	Asn	Leu	Gln	Asn 585	Ile	Pro	Ile	Arg	Thr 590	Pro	Leu	
GGC	CAG	AGG	ATC	CGC	CGG	GCC	TTC	GTG	GCC	GAG (	GCG (	GGA	TGG (	GCG 1	TTG	1824
Gly	Gln	Arg 595	Ile	Arg	Arg	Ala	Phe 600	Val	Ala	Glu	Ala	Gly 605	Trp	Ala	Leu	
GTG	GCC	CTG	GAC	TAT	AGC	CAG	ATA	GAG	CTC	CGG (	GTC (	CTC	GCC (	CAC C	TC	1872
Val	Ala 610	Leu	Asp	Tyr	Ser	Gln 615	Ile	Glu	Leu	Aŗg	Val 620	Leu	Ala	His	Leu	
TCC	GGG	GAC	GAG	AAC	CTG .	ATC	AGG	GTC '	TTC	CAG (	GAG (	GGG A	AAG (	GAC A	TC	1920
Ser 625	Gly	Asp	Glu	Asn	Leu 630	Ile	Arg	Val	Phe	Gln 635	Glu	Gly	Lys	Asp	11e 640	

CAC	ACC	CAG	ACC	GCA	AGC	TGG	ATG	TTC	GGC	GTC	TCC	CCG	GAG	GCC	GTG	1968
His	Thr	Gln	Thr	Ala 645	Ser	Trp	Met	Phe	Gly 650		Ser	Pro	Glu	Ala 655	Val	
GAC	CCC	CTG	ATG	CGC	CGG	GCG	GCC	AAG	ACG	GTG	AAC	TTC	GGC	GTC	CTC	2016
Asp	Pro	Leu	Met 660	Arg	Arg	Ala	Ala	Lys 665		Val	Asn	Phe	Gly 670		Leu	,
TAC	GGC	ATG	TCC	GCC	CAT	AGG	CTC	TCC	CAG	GAG	CTT	GCC	ATC	CCC	TAC	2064
Tyr	Gly	Met 675	Ser	Ala	His	Arg	Leu 680		Gln	Glu	Leu	Ala 685		Pro	Tyr	
GAG	GAG	GCG	GTG	GCC	TTT	ATA	GAG	CGC	TAC	TTC	CAA	AGC	TTC	CCC	AAG	2112
Glu	Glu 690	Ala	Val	Ala	Phe	11e 695	Glu	Arg	Tyr	Phe	G1n 700		Phe	Pro	Lys	
GTG	CGG.	GCC	TGG	ATA	GAA	AAG	ACC	CTG	GAG	GAG	GGG	AGG	AAG	CGG	GGC	2160
Val 705	Arg	Ala	Trp	Ile	Glu 710	Lys	Thr	Leu	Glu	Glu 715	_	Arg	, Lys	Arg	720	
TAC	GTG	GAA	ACC	CTC	TTC	GGA	AGA	AGG	CGC	TAC	GTG	ccc	GAC	CTC	AAC	2208
Tyr	Val	Glu	Thr	Leu 725	Phe	Gly	Arg	Arg	Arg 730	Tyr	Val	Pro	Asp	Leu 735	Asn	
GCC	CGG	GTG	AAG	AGC	GTC	AGG	GAG	GCC	GCG	GAG	CGC	ATG	GCC	TTC	AAC	2256
Ala	Arg	Val	Lys 740	Ser	Val	Arg	Glu	Ala 745	Ala	Glu	Arg	Met	Ala 750		Asn	
ATG	ccc	GTC	CAG	GGC	ACC	GCC	GCC	GAC	CTC	ATG	AAG	CTC	GCC	ATG	GTG	2304
Met	Pro	Val 755	Gln	Gly	Thr	Ala	Ala 760	Asp	Leu	Met	Lys	Leu 765		Met	Val	
AAG	CTC	TTC	ccc	CAC	CTC	CGG	GAG	ATG	GGG	GCC	CGC	ATG	CTC	CTC	CAG	2352
Lys	Leu 770	Phe	Pro	His		Arg 775		Met	Gly	Ala	Arg 780		Leu	Leu	Gln	
GTC	CAC	GAC	GAG	CTC	CTC	CTG	GAG	GCC	CCC	CAA	GCG	CGG	GCC	GAG	GAG	2400
Val 785	His	Asp	Glu	Leu	Leu 790	Leu	Glu	Ala	Pro	Gln 795	Ala	Arg	Ala	Glu	Glu 800	
GTG	GCG	GCT	TTG	GCC	AAG	GAG	GCC	ATG	GAG	AAG	GCC	TAT	ccc	CTC	GCC	2448
Val	Ala	Ala	Leu	Ala 805	Lys	Glu	Ala	Met	Glu 810	Lys	Ala	Tyr	Pro	Leu 815		

GTG CCC CTG GAG GTG GAG GTG GGG ATC GGG GAG GAC TGG CTT TCC GCC

Val Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala
820

825

830

AAG GGC TGA

2505

Lys Gly

# (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 834 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Lys Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1 5 10 15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly 20 25 30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala 35 40 45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe
50 55 60

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu 65 70 75 80

Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln 85 90 95

Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu 100 105 110

Glu Val Pro Gly Phe Glu Ala Asp Asp Val Leu Ala Thr Leu 1.1a Lys
115 120 125

Lys Ala Glu Arg Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg

Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu
145 150 155 160

Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Lys 165 170 175

Pro	Glu	Gln	180	Val	. Asp	Phe	Arg	Ala 185		Val	Gly	Asp	Pro 190		Asp
Asn	Leu	Pro 195	Gly	Val	. Lys	Gly	Ile 200		Glu	Lys	Thr	Ala 205		Lys	Leu
Leu	Lys 210		Trp	Gly	Ser	Leu 215		Asn	Ile	Leu	Lys 220		ı Leu	Asp	Arg
Val 225	Lys	Pro	Glu	Ser	Val 230		Glu	Arg	Ile	Lys 235		His	Leu	Glu	Asp 240
Leu	Lys	Leu	Ser	Leu 245	Glu	Leu	Ser	Arg	Val 250		Ser	Asp	Leu	Pro 255	Leu
Glu	Val	Asp	Phe 260	Ala	Arg	Arg	Arg	Glu 265	Pro	Asp	Arg	Glu	Gly 270		Arg
Ala	Phe	Leu 275	Glu	Arg	Leu	Glu	Phe 280	Gly	Ser	Leu	Leu	His 285		Phe	Gly
Leu	Leu 290	Glu	Ala	Pro	Ala	Pro 295	Leu	Glu	Glu	Ala	Pro 300	-	Pro	Pro	Pro
G1u 305	Gly	Ala	Phe	Val	Gly 310	Phe	Val	Leu	Ser	Arg 315	Pro	Glu	Pro	Met	Trp 320
				325		Ala			330					335	
Ala	Lys	Asp	Pro 340	Leu	Ala	Gly	Leu	Lys 345	Asp	Leu	Lys	Glu	Val 350	Arg	Gly
Leu	Leu	Ala 355	Lys	Asp	Leu	Ala	Val 360	Leu	Ala	Leu	Arg	Glu 365	Gly	Leu	Asp
Leu	Ala 370	Pro	Ser	Asp	Asp	Pro 375	Met	Leu	Leu	Ala	Tyr 380	Leu	Leu	Asp	Pro
Ser 385	Asn	Thr	Thr	Pro	Glu 390	Gly	Val	Ala <sub>.</sub>	Arg	Arg 395	Tyr	Gly	Gly	Glu	Trp 400
				405		Arg			410					415	
			420			Lys		425					430		-
Gln	Glu	Va1 435	Glu	Lys	Pro	Leu	Ser 440	Arg	Val	Leu	Ala	His 445	Met	Glu	Ala
Thr	Gly 450	Val	Arg	Leu	Asp	Val 455	Ala	Tyr	Leu		Ala 460	Leu	Ser	Leu	Glu
Leu 465	Ala	Glu	Glu	Ile	Arg 470	Arg	Leu	Glu		Glu 475	Val	Phe	Arg	Leu	Ala 480

Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu
485 490 495

Phe Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly 500 505 510

Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His 515 520 525

Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys 530 535 540

Asn Thr Tyr Val Asp Pro Leu Pro Gly Leu Val His Pro Arg Thr Gly 545 550 560

Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu 565 570 575

Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Ile Arg Thr Pro Leu 580 585 590

Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu 595 600 605

Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu 610 615 620

Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile 625 630 635 640

His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Ser Pro Glu Ala Val
645 650 655

Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu 660 665 670

Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr 675 680 685

Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys 690 695 700

Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly 705 710 715 720

Tyr Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn 725 730 735

Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn 740 745 750

Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val 755 760 765

Lys Leu Phe Pro His Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln 770 780

48

Val His Asp Glu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu 785 790 795 800

Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala 805 810 815

Val Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala 820 825 830

Lys Gly

#### (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2505 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Thermus thermophilus
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..2502
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
- ATG GAG GCG ATG CTT CCG CTC TTT GAA CCC AAA GGC CGG GTC CTC CTG

  Met Glu Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu

1 5 10 15

GTG GAC GGC CAC CTG GCC TAC CGC ACC TTC TTC GCC CTG AAG GGC 96

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly
20 25 30

CTC ACC ACG AGC CGG GGC GAA CCG GTG CAG GCG GTC TAC GGC TTC GCC 144

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala 35 40 45

AAG AGC CTC CTC AAG GCC CTG AAG GAG GAC GGG TAC AAG GCC GTC TTC 192

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe 50 60

-141-

GTG	GTC	TTT	GAC	GCC	AAG	GCC	CCC	TCC	TTC	CGC	CAC	GAG	GCC	TAC	GAG	240
Val 65		Phe	Asp	Ala	Lys 70	Ala	Pro	Ser	Phe	Arg 75		s Glu	ı Ala	а Туг	61u 80	
GCC	TAC	AAG	GCG	GGG	AGG	GCC	CCG	ACC	CCC	GAG	GAC	TTC	CCC	CGG	CAG	. 288
Ala	Tyr	Lys	Ala	G1y 85	_	Ala	Pro	Thr	Pro 90		Asp	Phe	Pro	Arg 95	Gln	
CTC	GCC	CTC	ATC	AAG	GAG	CTG	GTG	GAC	CTC	CTG	GGG	TTT	ACC	CGC	CTC	336
Leu	Ala	Leu	11e		Glu	Leu	Val	Asp 105	Leu	Leu	Gly	Phe	Thr 110		Leu	
GAG	GTC	CCC	GGC	TAC	GAG	GCG	GAC	GAC	GTT	CTC	GCC	ACC	CTG	GCC	AAG	384
Glu	Val	Pro 115		Tyr	Glu	Ala	Asp 120		Val	Leu	Ala	Thr 125		ı Ala	Lys	
AAG	GCG	GAA	AAG	GAG	GGG	TAC	GAG	GTG	CGC	ATC	CTC	ACC	GCC	GAC	CGC	432
Lys	Ala 130	Glu	Lys	Glu	Gly	Tyr 135	Glu	Val	Arg	Ile	Leu 140		Ala	Asp	Arg	
GAC	CTC	TAC	CAA	CTC	GTC	TCC	GAC	CGC	GTC	GCC	GTC	CTC	CAC	CCC	GAG	480
Asp 145	Leu	Tyr	Gln	Leu	Val 150	Ser	Asp	Arg	Val	Ala 155	Val	Leu	His	Pro	Glu 160	
GGC	CAC	CTC	ATC	ACC	CCG	GAG	TGG	CTT	TGG	GAG	AAG	TAC	GGC	CTC	AGG	528
Gly	His	Leu	Ile	Thr 165	Pro	Glu	Trp	Leu	Trp 170	Glu	Lys	Tyr	Gly	Leu 175	Arg	
CCG	GAG	CAG	TGG	GTG	GAC	TTC	CGC	GCC	CTC	GTG	GGG	GAC	CCC	TCC	GAC	576
Pro	Glu	Gln	Trp 180	Val	Asp	Phe	Arg	Ala 185	Leu	Val	Gly	Asp	Pro 190		Asp	
AAC	CTC	CCC	GGG	GTC	AAG	GGC	ATC	GGG	GAG .	AAG .	ACC	GCC	CTC	AAG	CTC	624
Asn	Leu	Pro 195	Gly	Val	Lys	Gly	11e 200	Gly	Glu	Lys	Thr	Ala 205	Leu	Lys	Leu	
стс	AAG	GAG	TGG	GGA	AGC	CTG	GAA	AAC	CTC	CTC .	AAG .	AAC	CTG	GAC (	CGG	672
	Lys 210	Glu	Trp	Gly	Ser	Leu 215	Glu	Asn	Leu	Leu	Lys 220	Asn	Leu	Asp	Arg	
GTA	AAG	CCA	GAA	AAC	GTC	CGG	GAG .	AAG .	ATC A	AAG (	GCC	CAC	CTG	GAA (	GAC	720
Val 225	Lys	Pro	Glu	Asn	Val 230	Arg	Glu	Lys	Ile	Lys 235	Ala	His	Leu	Clu	Asp 240	

CTC	AGG	CTC	TCC	TTG	GAG	CTC	TCC	CGG	GTG	CGC	ACC	GAC	CTC	CCC	CTG	768	
Leu	Arg	Leu	Ser	Leu 245		Leu	Ser	Arg	Val 250		; Thr	Asp	Leu	255	Leu .		
GAG	GTG	GAC	CTC	GCC	CAG	GGG	CGG	GAG	ССС	GAC	CGG	GAG	GGG	CTT	AGG	816	
Glu	Val	Asp	Leu 260	Ala	Gln	Gly	Arg	Glu 265	Pro	Asp	Arg	G1u	G1y 270		Arg		
GCC	TTC	CTG	GAG	AGG	CTG	GAG	TTC	GGC	AGC	CTC	CTC	CAC	GAG	TTC	GGC	864	-
Ala	Phe	Leu 275	Glu	Arg	Leu	Glu	Phe 280	Gly	Ser	Leu	Leu	His 285		Phe	Gly		
CTC	CTG	GAG	GCC	CCC	GCC	CCC	CTG	GAG	GAG	GCC	ССС	TGG	CCC	CCG	CCG	912	
Leu	Leu 290	Glu	Ala	Pro	Ala	Pro 295	Leu	Glu	Glu	Ala	Pro 300		Pro	Pro	Pro		
GAA	GGG	GCC	TTC	GTG	GGC	TTC	GTC	CTC	TCC	CGC	CCC	GAG	CCC	ATG	TGG	960	
Glu 305	Gly	Ala	Phe	Val	Gly 310	Phe	Val	Leu	Ser	Arg 315	Pro	Glu	Pro	Met	Trp 320		
GCG	GAG	CTT	AAA	GCC	CTG	GCC	GCC	TGC	AGG	GAC	GGC	CGG	GTG	CAC	CGG	1008	, .
Ala	Glu	Leu	Lys	Ala 325	Leu	Ala	Ala	Cys	Arg 330	Asp	Gly	Arg	Val	His 335	Arg		
GCA	GCA	GAC	CCC	TTG	GCG	GGG	CTA	AAG	GAC	стс	AAG	GAG	GTC	CGG	GGC	1056	
Ala	Ala	Asp	Pro 340	Leu	Ala	Gly	Leu	Lys 345	Asp	Leu	Lys	Glu	Val 350		Gly		
CTC	CTC	GCC	AAG	GAC	CTC	GCC	GTC	TTG	GCC '	TCG	AGG	GAG	GGG	CTA (	GAC	1104	
Leu	Leu	Ala 355	Lys	Asp	Leu	Ala	Val 360	Leu	Ala	Ser	Arg	Glu 365	Gly	Leu	Asp		
CTC	GTG	CCC	GGG	GAC	GAC	ccc .	ATG	CTC (	CTC (	GCC	TAC	CTC	CTG (	GAC (	ccc	1152	
Leu	Val 370	Pro	Gly	Asp	Asp	Pro 375	Met	Leu	Leu	Ala	Tyr 380	Leu	Leu	Asp	Pro		
TCC	AAC	ACC .	ACC	CCC	GAG	GGG (	GTG (	GCG (	CGG (	cgc '	TAC (	GGC (	GGG (	GAG :	rgg	1200	
Ser 385	Asn	Thr	Thr	Pro	Glu 390	Gly	Val	Ala	Arg	Arg 395	Tyr	Gly	Gly	Glu	Trp 400		
ACG	GAG	GAC (	GCC (	GCC	CAC	CGG (	GCC (	CTC (	CTC 1	rcg (	GAG A	AGG (	CTC (	CAT (	CGG	1248	
Thr	Glu .	Asp .		Ala 405	His	Arg	Ala		Leu 410	Ser	Glu	Arg	Leu	llis	Arg		

AAC	CTO	CTI	. AAG	CGC	стс	GAG	GGG	GAG	GAG	AAC	CTC	CTI	TGG	CTC	TAC	129
Asn	Let	ı Leı	Lys 420	_	, Leu	ı Glu	Gly	7 Gla 425		u Ly	s Le	u Le	u Tr 43		u Tyr	
CAC	GAG	GTG	GAA	AAG	CCC	CTC	TCC	CGG	GTC	CTG	GCC	CAC	ATG	GAG	GCC	1344
His	Glu	Val 435		Lys	Pro	Leu	Ser 440		g Val	L Le	u Al	a Hi 44		t Glu	ı Ala	
ACC	GGG	GTA	CGG	CTG	GAC	GTG	GCC	TAC	CTT	CAG	GCC	CTT	TCC	CTG	GAG	1392
Thr	G1y 450		Arg	Leu	Asp	Val 455	Ala	Tyr	Leu	ı Glı	n Ala 460		u Sei	r Lev	ı Glu	
CTT	GCG	GAG	GAG	ATC	CGC	CGC	CTC	GAG	GAG	GAG	GTC	TTC	CGC	TTG	GCG	1440
Leu 465	Ala	Glu	Glu	Ile	Arg 470	_	Leu	Glu	Glu	Glu 475	_	l Phe	e Arg	g Leu	Ala 480	
GGC	CAC	ccc	TTC	AAC	CTC	AAC	TCC	CGG	GAC	CAG	CTG	GAA	AGG	GTG	CTC	1488
Gly	His	Pro	Phe	Asn 485	Leu	Asn	Ser	Arg	Asp 490		. Leu	ı Glu	ı Arg	Val 495		
TTT	GAC	GAG	CTT	AGG	CTT	ССС	GCC	TTG	GGG	AAG	ACG	CAA	AAG	ACA	GGC	1536
Phe	Asp	Glu	Leu 500	Arg	Leu	Pro	Ala	<b>Leu</b> 505	Gly	Lys	Thr	Gln	Lys 510		Gly	
AAG	CGC	TCC	ACC	AGC	GCC	GCG	GTG	CTG	GAG	GCC	CTA	CGG	GAG	GCC (	CAC	1584
Lys	Arg	Ser 515	Thr	Ser	Ala	Ala	Val 520	Leu	Glu	Ala	Leu	Arg 525		Ala	His	
CCC	ATC	GTG	GAG .	AAG .	ATC	CTC (	CAG	CAC	CGG	GAG	CTC	ACC	AAG	CTC A	AAG	1632
	Ile 530	Val	Glu	Lys	Ile	Leu 535	Gln	His	Arg	Glu	Leu 540	Thr	Lys	Leu	Lys	
AAC	ACC	TAC	GTG (	GAC	ccc	CTC (	CCA A	AGC	CTC	GTC	CAC	CCG .	AGG A	ACG C	GC	1680
Asn 545	Thr	Tyr	Val .		Pro 550	Leu	Pro	Ser	Leu	Val 555	His	Pro	Arg	Thr	Gly 560	
CGC	CTC	CAC	ACC (	CGC :	TTC A	AAC (	CAG A	ACG (	GCC A	ACG	GCC A	ACG (	GGG A	AGG C	TT	1728
Arg :	Leu	His		Arg 565	Phe	Asn	Gln	Thr	Ala 570	Thr	Ala	Thr	Gly	Arg 575	Leu	
AGT A	AGC	TCC	GAC (	CCC A	AAC (	CTG C	AG A	AC A	ATC (	CCC (	GTC (	CGC A	ACC (	CCC T	TG	1776
Ser :	Ser		Asp 1	Pro A	Asn	Leu (		Asn 585	Ile	Pro	Val	Arg	Thr 590	Fro	Leu	

GGC	CAG	AGG	ATC	CGC	CGG	GCC	TTC	GTG	GCC	GAG	GCG	GGT	TGG	GCG	TTG	1824	
Gly	Gln	Arg 595		Arg	Arg	Ala	Phe 600		. Ala	Glu	ı Al	a Gl; 60		p Al	a Leu		
GTG	GCC	CTG	GAC	TAT	AGC	CAG	ATA	GAG	CTC	CGC	GTC	CTC	GCC	CAC	CTC	1872	
Val	Ala 610	Leu	Asp	Tyr	Ser	Gln 615		Glu	Leu	Arg	y Va 62	_	u Ala	a His	s Leu		
TCC	GGG	GAC	GAA	AAC	CTG	ATC	AGG	GTC	TTC	CAG	GAG	GGG	AAG	GAC	ATC	1920	
Ser 625	Gly	Asp	Glu	Asn	Leu 630		Arg	Val	Phe	Glr 635		u Gly	y Lys	s Ası	0 Ile 640		
CAC	ACC	CAG	ACC	GCA	AGC	TGG	ATG	TTC	GGC	GTC	CCC	CCG	GAG	GCC	GTG	1968	
His	Thr	Gln	Thr	Ala 645	Ser	Trp	Met	Phe	Gly 650		Pr	Pro	Glu	1 Ala 655	a Val		
GAC	CCC	CTG	ATG	CGC	CGG	GCG	GCC	AAG	ACG	GTG	AAC	TTC	GGC	GTC	CTC	2016	
Asp	Pro	Leu	Met 660	Arg	Arg	Ala	Ala	Lys 665	Thr	Val	Ası	n Phe	Gly 670		Leu		
TAC	GGC	ATG	TCC	GCC	CAT	AGG	CTC	TCC	CAG	GAG	CTT	GCC	ATC	ccc	TAC	2064	
Tyr	Gly	Met 675	Ser	Ala	His	Arg	Leu 680	Ser	Gln	Glu	Let	1 Ala 685		Pro	Tyr		
GAG	GAG	GCG	GTG	GCC	TTT	ATA	GAG	CGC	TAC	TTC	CAA	AGC	TTC	ССС	AAG	2112	
Glu	Glu 690	Ala	Val	Ala	Phe	Ile 695	Glu	Arg	Tyr	Phe	Gl <sub>1</sub> 700		Phe	Pro	Lys		
GTG	CGG	GCC	TGG	ATA	GAA	AAG	ACC	CTG	GAG	GAG	GGG	AGG	AAG	CGG	GGC	2160	
Val 705	Arg	Ala	Trp	Ile	Glu 710	Lys	Thr	Leu	Glu	Glu 715	Gly	Arg	, Lys	Arg	Gly 720		
TAC	GTG	GAA	ACC	CTC	TTC	GGA .	AGA .	AGG	CGC	TAC	GTG	ccc	GAC	CTC	AAC	2208	
Tyr	Val	Glu	Thr	Leu 725	Phe	Gly	Arg	Arg	Arg 730	Tyr	Val	. Pro	Asp	Leu 735	Asn		
GCC	CGG	GTG	AAG .	AGC	GTC .	AGG	GAG	GCC	GCG	GAG	CGC	ATG	GCC	TTC	AAC	2256	
Ala	Arg	Val	Lys 740	Ser	Val	Arg	Glu	Ala 745	Ala	Glu	Arg	, Met	Ala 750		Asn		
ATG	CCC	GTC	CAG	GGC .	ACC (	GCC 4	GCC (	GAC	CTC .	ATG .	AAG	CTC	GCC .	ATG	GTG	2304	
Met	Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val		

AAG	CTC	TTC	CCC	CGC	CTC	CGG	GAG	ATG	GGG	GCC	CGC	ATG	CTC	CTC (	CAG	2352
Lys	Leu 770		Pro	Arg	Leu	Arg 775		Met	Gly	Ala	Arg 780		Leu	Leu	Gln	
GTC	CAC	GAC	GAG	CTC	CTC	CTG	GAG	GCC	CCC	CAA	GCG	CGG	GCC	GAG C	GAG	2400
Val 785	His	Asp	Glu	Leu	Leu 790	Leu	Glu	Ala	Pro	G1n 795		Arg	Ala	Glu	Glu 800	
GTG	GCG	GCT	TTG	GCC	AAG	GAG	GCC	ATG	GAG	AAG	GCC	TAT	CCC (	CTC G	CC	2448
Val	Ala	Ala	Leu	Ala 805	Lys	Glu	Ala	Met	Glu 810	Lys	Ala	Tyr	Pro	Leu 815	Ala	•
GTG	ccc	CTG	GAG	GTG	GAG	GTG	GGG	ATG	GGG	GAG	GAC	TGG (	CTT 1	CC G	CC	2496
Val	Pro	Leu	Glu 820	Val	Glu	Val	Gly	Met 825	Gly	Glu	Asp	Trp	Leu 830	Ser	Ala	
AAG	GGT	TAG														2505
Lys	Gly															

#### (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 834 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Glu Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu 1 5 10 15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly 20 25 30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala 35 40 45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe 50 55 60

Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu 65 70 75 80

Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
85 90 95

Le	ı Ala	a Le	u I1 10	e Ly 0	's Gl	u Le	u Va	1 As	p Lei 5	u Le	u Gl	y Ph	e Th		g Leu
Glı	ı Val	l Pr 11	o Gl 5	у Ту	r Gl	u Ala	a As <sub>1</sub>	p As <sub>]</sub>	p Va	l Le	u Al	a Th 12		u Al	a Lys
Lys	130	a G1:	u Ly	s Gl	u Gl	y Ty:	r Glu	u Va	l Ar	g I16	Lei 140		r Al	a As	p Arg
Asp 145	Leu	Ту	r Gl:	n Le	u Va 150	l Sei	c Ası	Arg	g Va]	l Ala 155		l Le	u Hi	s Pr	o Glu 160
Gly	His	Let	ı 11e	e Th	r Pro	o Glu	ı Trş	Let	17rp	Glu	ı Lys	з Туг	r Gly	y Le	u Arg 5
Pro	Glu	Glr	180	o Vai	l As <sub>l</sub>	Phe	e Arg	3 Ala 185	Leu i	Val	. Gly	/ Ası	Pro 190		r Asp
Asn	Leu	Pro 195	Gly	y Val	l Lys	Gly	7 Ile 200	Gly	Glu	Lys	Thr	Ala 205		ı Ly:	s Leu
Leu	Lys 210	<b>Gl</b> u	Trp	Gly	y Ser	Leu 215	Glu	. Asn	Leu	Leu	Lys 220		Lei	ı Asp	Arg
Val 225	Lys	Pro	Glu	ı Ası	230	Arg	Glu	Lys	Ile	Lys 235	Ala	His	Leu	ı Glu	Asp 240
				243	)				250					255	•
			260	ı				265					270	l	Arg
		2/3				Glu	280					285			
	290					Pro 295					300				
303					310	Phe				315					320
				325		Ala			330					335	
Ala	Ala	Asp	Pro 340	Leu	Ala	Gly	Leu	Lys 345	Asp	Leu	Lys	Glu	Val 350	Arg	Gly
Leu	Leu	Ala 355	Lys	Asp	Leu	Ala	Val 360	Leu	Ala	Ser	Arg	Glu 365	Gly	Leu	Asp
Leu	Val 370	Pro	Gly	Asp	Asp	Pro 375	Met	Leu	Leu		Tyr 380	Leu	Leu	Asp	Pro
Ser 385	Asn	Thr	Thr	Pro	Glu 390	Gly	Val	Ala		Arg 395	Tyr	Gly	Gly	Glu	Trp 400

Thr Glu Asp Ala Ala His Arg Ala Leu Leu Ser Glu Arg Leu His Arg
405 410 415

Asn Leu Leu Lys Arg Leu Glu Gly Glu Glu Lys Leu Leu Trp Leu Tyr
420 425 430

His Glu Val Glu Lys Pro Leu Ser Arg Val Leu Ala His Met Glu Ala 435 440 445

Thr Gly Val Arg Leu Asp Val Ala Tyr Leu Gln Ala Leu Ser Leu Glu
450 460

Leu Ala Glu Glu Ile Arg Arg Leu Glu Glu Glu Val Phe Arg Leu Ala 465 470 475 480

Gly His Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg <sup>9</sup>al Leu 485 490 495

Phe Asp Glu Leu Arg Leu Pro Ala Leu Gly Lys Thr Gln Lys Thr Gly 500 505 510

Lys Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His 515 520 525

Pro Ile Val Glu Lys Ile Leu Gln His Arg Glu Leu Thr Lys Leu Lys 530 535 540

Asn Thr Tyr Val Asp Pro Leu Pro Ser Leu Val His Pro Arg Thr Gly 545 550 555 560

Arg Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu 565 570 575

Ser Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu
580 585 590

Gly Gln Arg Ile Arg Arg Ala Phe Val Ala Glu Ala Gly Trp Ala Leu 595 600 605

Val Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu 610 615 620

Ser Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile 625 630 635 640

His Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val
645 650 655

Asp Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu 660 665 670

Tyr Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr 675 680 685

Glu Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys 690 700

Val Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly 705 710 715 720

Tyr Val Glu Thr Leu Phe Gly Arg Arg Tyr Val Pro Asp Leu Asn 725 730 735

Ala Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn 740 745 750

Met Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val 755 760 765

Lys Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln
770 780

Val His Asp Glu Leu Leu Clu Ala Pro Gln Ala Arg Ala Glu Glu 785 790 795 800

Val Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala 805 810 315

Val Pro Leu Glu Val Glu Val Gly Met Gly Glu Asp Trp Leu Ser Ala 820 825 830

Lys Gly

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2679 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:

(A) ORGANISM: Thermosipho africanus

- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..2676
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATG GGA AAG ATG TTT CTA TTT GAT GGA ACT GGA TTA GTA TAC AGA GCA

Met Gly Lys Met Phe Leu Phe Asp Gly Thr Gly Leu Val Tyr Arg Ala 1 5 10 15

TTT	TAT	GCT	ATA	GAT	CAA	TCT	CTT	CAA	ACT	TCG	TCT	GGT	TTA	CAC	ACT	96
Phe	Tyr	Ala	Ile 20	Asp	Gln	Ser	Leu	G1n 25	Thr	Ser	Ser	Gly	Leu 30	His	Thr	
AAT	GCT	GTA	TAC	GGA	CTT	ACT	AAA	ATG	CTT	ATA	AAA	TTT	TTA	AAA	GAA	144
Asn	Ala	Val 35	Tyr	Gly	Leu	Thr	Lys 40	Met	Leu	Ile	Lys	Phe 45	Leu	Lys	Glu	
CAT	ATC	AGT	ATT	GGA	AAA	GAT	GCT	TGT	GTT	TTT	GTT	TTA	GAT	TCA	AAA	192
His	Ile 50	Ser	Ile	Gly	Lys	Asp 55	Ala	Cys	Val	Phe	Val 60	Leu	Asp	Ser	Lys	
GGT	GGT	AGC	AAA	AAA	AGA	AAG	GAT	ATT	CTT	GAA	ACA	TAT	AAA	GCA	AAT	240
G1y 65	Gly	Ser	Lys	Lys	Arg 70	Lys	Asp	Ile	Leu	Glu 75	Thr	Tyr	Lys	Ala	Asn 80	
AGG	CCA	TCA	ACG	CCT	GAT	TTA	CTT	TTA	GAG	CAA	ATT	CCA	TAT	GTA	GAA	288
Arg	Pro	Ser	Thr	Pro 85	Asp	Leu	Leu	Leu	Glu 90	Gln	Ile	Pro	Tyr	Val 95	Glu	
GAA	CTT	GTT	GAT	GCT	CTT	GGA	ATA	AAA	GTT	TTA	AAA	ATA	GAA	GGC	TTT	336
Glu	Leu	Val	Asp 100	Ala	Leu	Gly	Ile	Lys 105		Leu	Lys	Ile	Glu 110	Gly	Phe	
GAA	GCT	GAT	GAC	ATT	ATT	GCT	ACG	CTT	TCT	AAA	AAA	TTT	GAA	AGT	GAT	384
Glu	Ala	Asp 115	Asp	Ile	Ile	Ala	Thr 120	Leu	Ser	Lys	Lys	Phe 125	Glu	Ser	Asp	
TTT	GAA	AAG	GTA	AAC	ATA	ATA	ACT	GGA	GAT	AAA	GAT	CTT	TTA	CAA	CTT	432
Phe	Glu 130	Lys	Val	Asn	Ile	Ile 135	Thr	Gly	Asp	Lys	Asp 140		Leu	Cln	Leu	
GTT	TCT	GAT	AAG	GTT	TTT	GTT	TGG	AGA	GTA	GAA	AGA	GGA	ATA	ACA	GAT	480
Val 145	Ser	Asp	Lys	Val	Phe 150	Val	Trp	Arg	Val	Glu 155	Arg	G1y	Ile	Thr	Asp 160	
TTG	GTA	TTG	TAC	GAT	AGA	AAT	AAA	GTG	ATT	GAA	AAA	TAT	GGA	ATC	TAC	528
Leu	Val	Leu	Tyr	Asp 165	Arg	Asn	Lys	Val	Ile 170	Glu	Lys	Tyr	Gly	Ile 175	Tyr	
CCA	GAA	CAA	TTC	AAA	GAT	TAT	ATT	TCT	CTT	GTC	GGT	GAT	CAG	ATT	GAT	576
Pro	Glu	Gln	Phe 180	Lys	Asp	Tyr	Leu	Ser 185	Leu	Val	Gly	Asp	Gln 190	Ile	Asp	

AAT ATC CCA GGA GTT AAA GGA ATA GGA AAG AAA ACA GCT GTT TCG C	TT 624
Asn Ile Pro Gly Val Lys Gly Ile Gly Lys Lys Thr Ala Val Ser 195 200 205	Leu
TTG AAA AAA TAT AAT AGC TTG GAA AAT GTA TTA AAA AAT ATT AAC C	TT 672
Leu Lys Lys Tyr Asn Ser Leu Glu Asn Val Leu Lys Asn Ile Asn 210 215 220	Leu
TTG ACG GAA AAA TTA AGA AGG CTT TTG GAA GAT TCA AAG GAA GAT T	TG 720
Leu Thr Glu Lys Leu Arg Arg Leu Leu Glu Asp Ser Lys Glu Asp 225 230 235	Leu 240
CAA AAA AGT ATA GAA CTT GTG GAG TTG ATA TAT GAT GTA CCA ATG G	AT 768
Gln Lys Ser Ile Glu Leu Val Glu Leu Ile Tyr Asp Val Pro Met 245 250 255	Asp
GTG GAA AAA GAT GAA ATA ATT TAT AGA GGG TAT AAT CCA GAT AAG C	TT. 816
Val Glu Lys Asp Glu Ile Ile Tyr Arg Gly Tyr Asn Pro Asp Lys 260 265 270	Leu
TTA AAG GTA TTA AAA AAG TAC GAA TTT TCA TCT ATA ATT AAG GAG T	TA 864
Leu Lys Val Leu Lys Lys Tyr Glu Phe Ser Ser Ile Ile Lys Glu 275 280 285	Leu
AAT TTA CAA GAA AAA TTA GAA AAG GAA TAT ATA CTG GTA GAT AAT G	AA 912
Asn Leu Gln Glu Lys Leu Glu Lys Glu Tyr Ile Leu Val Asp Asn 290 295 300	Glu
GAT AAA TTG AAA AAA CTT GCA GAA GAG ATA GAA AAA TAC AAA ACT T	TT 960
Asp Lys Leu Lys Lys Leu Ala Glu Glu Ile Glu Lys Tyr Lys Thr 305 310 315	Phe 320
TCA ATT GAT ACG GAA ACA ACT TCA CTT GAT CCA TTT GAA GCT AAA C	TG 1008
Ser Ile Asp Thr Glu Thr Thr Ser Leu Asp Pro Phe Glu Ala Lys 325 330 335	Leu
GTT GGG ATC TCT ATT TCC ACA ATG GAA GGG AAG GCG TAT TAT ATT C	CG 1056
Val Gly Ile Ser Ile Ser Thr Met Glu Gly Lys Ala Tyr Tyr Ile 340 345 350	Pro
GTG TCT CAT TTT GGA GCT AAG AAT ATT TCC AAA AGT TTA ATA GAT A	AA 1104
Val Ser His Phe Gly Ala Lys Asn Ile Ser Lys Ser Leu Ile Asp 355 360 365	Lys

TTT	CTA	AAA	CAA	ATT	TTG	CAA	GAG	AAG	GAT	TAT	AAT	ATC	GTT	GGT	CAG	1152
Phe	Leu 370	Lys	Gln	Ile	Leu	Gln 375	Glu	Lys	Asp	Tyr	Asn 380	Ile	· Val	Gly	/ Gln	
AAT	TTA	AAA	TTT	GAC	TAT	GAG	ATT	TTT	AAA	AGC	ATG	GGT	TTT	TCT	CCA	1200
Asn 385	Leu	Lys	Phe	Asp	Tyr 390	Glu	Ile	Phe	Lys	Ser 395	Met	Gly	Phe	e Sei	400	
AAT	GTT	CCG	CAT	TTT	GAT	ACG	ATG	ATT	GCA	GCC	TAT	CTT	TTA	AAT	CCA	1248
Asn	Val	Pro	His	Phe 405	Asp	Thr	Met	Ile	Ala 410	Ala	Tyr	Leu	ı Lev	415	r Pro	
GAT	GAA	AAA	CGT	TTT	AAT	CTT	GAA	GAG	CTA	TCC	TTA	AAA	TAT	TTA	GGT	1296
Asp	Glu	Lys	Arg 420	Phe	Asn	Leu	Glu	Glu 425	Leu	Ser	Leu	Lys	430	Leu	ı Gly	
TAT	AAA	ATG	ATC	TCG	TTT	GAT	GAA	TTA	GTA	AAT	GAA	AAT	GTA	CCA	TTG	1344
Tyr	Lys	Met 435	Ile	Ser	Phe	Asp	Glu 440		Val	Asn	Glu	Asn 445		Pro	Leu	
TTT	GGA	AAT	GAC	TTT	TCG	TAT	GTT	CCA	CTA	GAA	AGA	GCC	GTT	GAG	TAT	1392
Phe	Gly 450	Asn	Asp	Phe	Ser	Tyr 455	Val	Pro	Leu	G1u	Arg 460	Ala	Val	. Glu	Tyr	
TCC	TGT	GAA	GAT	GCC	GAT	GTG	ACA	TAC	AGA	ATA	TTT	AGA	AAG	CTT	GGT	1440
Ser 465	Cys	Glu	Asp	Ala	Asp 470	Val	Thr	Tyr	Arg	Ile 475	Phe	Arg	Lys	: Lev	480	
AGG	AAG	ATA	TAT	GAA	AAT	GAG	ATG	GAA	AAG	TTG	TTT	TAC	GAA	ATT	GAG	1488
Arg	Lys	Ile	Tyr	G1u 485	Asn	Glu	Met	Glu	Lys 490	Leu	Phe	Tyr	Glu	495	Glu	
ATG	CCC	TTA	ATT	GAT	GTT	CTT	TCA	GAA	ATG	GAA	CTA	AAT	GGA	GTG	TAT	1536
Met	Pro	Leu	Ile 500	Asp	Val	Leu	Ser	G1u 505	Met	Glu	Leu	Asn	61y 510	Val	. Tyr	
TTT	GAT	GAG	GAA	TAT	TTA	AAA	GAA	TTA	TCA	AAA	AAA	TAT	CAA	GAA	AAA	1584
Phe	Asp	Glu 515	Glu	Tyr	Leu	Lys	Glu 520	Leu	Ser	Lys	Lys	Tyr 525	Gln	Glu	Lys	
ATG	GAT	GGA	ATT	AAG	GAA	AAA	GTT	TTT	GAG	ATA	GCT	GGT	GAA	ACT	TTC	1632
Met	Asp 530	Gly	Ile	Lys	Glu	Lys 535	Val	Phe	Glu	Ile	Ala 540	Gly	Glu	Thr	Phe	

AAT	TTA	AAC	TCT	TCA	ACT	CAA	GTA	GCA	TAT	ATA	CTA	TTT	GAA	AAA '	TTA	1680	
Asn 545	Leu	Asn	Ser	Ser	Thr 550		Val	Ala	Tyr	11e 555		Phe	e Glu	Lys	Leu 560		<u>.</u>
AAT	ATT	GCT	CCT	TAC	AAA	AAA	ACA	GCG	ACT	GGT	AAG	TTT	TCA	ACT A	AAT	1728	•
Asn	Ile	Ala	Pro	Tyr 565	Lys	Lys	Thr	Ala	Thr 570	-	Lys	Phe	Ser	Thr 575	Asn		á
GCG	GAA	GTT	TTA	GAA	GAA	CTT	TCA	AAA	GAA	CAT	GAA	ATT	GCA	AAA :	<b>IT</b> G	1776	
Ala	Glu	Val	Leu 580	Glu	Glu	Leu	Ser	Lys 585		His	Glu	Ile	Ala 590	Lys	Leu		
TTG	CTG	GAG	TAT	CGA	AAG	TAT	CAA	AAA	TTA	AAA	AGT	ACA	TAT	ATT (	GAT	1824	
Leu	Leu	Glu 595	Tyr	Arg	Lys	Tyr	Gln 600		Leu	Lys	Ser	Thr 605	-	Ile	Asp		
TCA	ATA	CCG	TTA	TCT	ATT	AAT	CGA	AAA	ACA	AAC	AGG	GTC	CAT	ACT A	ACT	1872	
Ser	Ile 610	Pro	Leu	Ser	Ile	Asn 615	Arg	Lys	Thr	Asn	Arg 620		His	Thr	Thr		
TTT	CAT	CAA	ACA	GGA	ACT	TCT	ACT	GGA	AGA	TTA	AGT	AGT	TCA	AAT (	CCA	1920	
Phe 625	His	G1n	Thr	Gly	Thr 630	Ser	Thr	Gly	Arg	Leu 635		Ser	Ser	Asn	Pro 640		
AAT	TTG	CAA	AAT	CTT	CCA	ACA	AGA	AGC	GAA	GAA	GGA	AAA	GAA	ATA A	\GA	1968	
Asn	Leu	Gln	Asn	Leu 645	Pro	Thr	Arg	Ser	Glu 650		Gly	Lys	Glu	Ile 655	Arg		
AAA	GCA	GTA	AGA	CCT	CAA	AGA	CAA	GAT	TGG	TGG	ATT	ATT	GGT	GCT (	GAC	2016	
Lys	Ala	Val	Arg 660	Pro	Gln	Arg	Gln	Asp 665		Trp	Ile	Leu	Gly 670	Ala	Asp		
TAT	TCT	CAG	ATA	GAA	CTA	AGG	GTT	ATT	GCG	CAT	GTA	AGT	AAA	GAT C	SAA	2064	
Tyr		Gln 675		Glu	Leu		Val 680		Ala	His	Val	Ser 685		Asp	Glu		
AAT	CTA	CTT	AAA	GCA	TTT	AAA	GAA	GAT	TTA	GAT	ATT	CAT	ACA .	ATT A	CT	2112	Ĩ
Asn	Leu 690	Leu	Lys	Ala	Phe	Lys 695	Glu	Asp	Leu	Asp	Ile 700	His	Thr	Ile	Thr		•
GCT	GCC	AAA	ATT	TTT	GGT	GTT	TCA	GAG	ATG	TTT	GTT	AGT	ĞAA	CAA A	ATG	2160	
Ala 705	Ala	Lys	Ile	Phe	Gly 710	Val	Ser	Glu	Met	Phe 715		Ser	Glu	Gln	Met 720		

AGA	AGA	GTT	GGA	AAG	ATG	GTA	AAT	TTT	GCA	ATT	ATT	TAT	GGA	GTT	TCA	2208
Arg	Arg	Val	Gly	Lys 725	Met	Val	Asn	Phe	Ala 730	Ile	Ile	Tyr	Gly	Val 735	Ser	
CCT	TAT	GGT	CTT	TCA	AAG	AGA	ATT	GGT	CTT	AGT	GTT	TCA	GAG	ACT	AAA	2256
Pro	Tyr	Gly	Leu 740	Ser	Lys	Arg	Ile	Gly 745	Leu	Ser	Val	Ser	G1u 750	Thr	Lys	
AAA	ATA	ATA	GAT	AAC	TAT	TTT	AGA	TAC	TAT	AAA	GGA	GTT	TTT	GAA	TAT	2304
Lys	Ile	11e 755	Asp	Asn	Tyr	Phe	Arg 760	Tyr	Tyr	Lys	Gly	Val 765	Phe	Glu	Tyr	
TTA	AAA	AGG	ATG	AAA	GAT	GAA	GCA	AGG	AAA	AAA	GGT	TAT	GTT	ACA	ACG	2352
Leu	Lys 770	Arg	Met	Lys	Asp	Glu 775	Ala	Arg	Lys	Lys	Gly 780	Tyr	Val	Thr	Thr	
CTT	TTT	GGA	AGG	CGC	AGA	TAT	ATT	CCA	CAG	TTA	AGA	TCG	AAA	AAT	GGT	2400
Leu 785	Phe	Gly	Arg	Arg	Arg 790	Tyr	Ile	Pro	Gln	Leu 795	Arg	Ser	Lys	Asn	Gly 800	
AAT	AGA	GTT	CAA	GAA	GGA	GAA	AGA	ATA	GCT	GTA	AAC	ACT	CCA	ATT	CAA	2448
Asn	Arg	Val	Gln	Glu 805	Gly	Glu	Arg	Ile	Ala 810	Val	Asn	Thr	Pro	11e 815	Gln	
GGA	ACA	GCA	GCT	GAT	ATA	ATA	AAG	ATA	GCT	ATG	ATT	AAT	ATT	CAT	AAT	2496
Gly	Thr	Ala	Ala 820	Asp	Ile	Ile	Lys	Ile 825	Ala	Met	Ile	Asn	11e 830	Pis	Asn	
AGA	TTG	AAG	AAG	GAA	AAT	CTA	CGT	TCA	AAA	ATG	ATA	TTG	CAG	GTT	CAT	2544
Arg	Leu	Lys 835	Lys	Glu	Asn	Leu	Arg 840		Lys	Met	Ile	Leu 845	Gln	∖'al	His	
GAC	GAG	TTA	GTT	TTT	GAA	GTG	CCC	GAT	AAT	GAA	CTG	GAG	ATT	GTA .	AAA	2592
Asp	Glu 850	Leu	Val	Phe	Glu	Val 855	Pro	Asp	Asn	Glu	Leu 860	Glu	Ile	Val	Lys	
GAT	TTA	GTA	AGA	GAT	GAG	ATG	GAA	TAA	GCA	GTT	AAG	CTA	GAC	GTT	CCT	2640
Asp 865	Leu	Val	Arg	Asp	Glu 870	Met	Glu	Asn	Ala	Val 875	Lys	Leu	Asp	Val	Pro 880	
TTA	AAA	GTA	GAT	GTT	TAT	TAT	GGA	AAA	GAG	TGG	GAA	TAA	٠			2679
Leu	Lys	Val	Asp	Val 885	Tyr	Tyr	Gly	Lys	Glu 890	Trp	Glu					

(2) INFORMATIO	n for	SEQ	ID	NO:	12	:
----------------	-------	-----	----	-----	----	---

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 892 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Gly Lys Met Phe Leu Phe Asp Gly Thr Gly Leu Val Tyr Arg Ala 1 5 10 15

Phe Tyr Ala Ile Asp Gln Ser Leu Gln Thr Ser Ser Gly Leu His Thr 20 25 30

Asn Ala Val Tyr Gly Leu Thr Lys Met Leu Ile Lys Phe Leu Lys Glu 35 40 45

His Ile Ser Ile Gly Lys Asp Ala Cys Val Phe Val Leu Asp Ser Lys 50 55 60

Gly Gly Ser Lys Lys Arg Lys Asp Ile Leu Glu Thr Tyr Lys Ala Asn 65 70 75 80

Arg Pro Ser Thr Pro Asp Leu Leu Glu Glu Gln Ile Pro Tyr Val Glu 85 90 95

Glu Leu Val Asp Ala Leu Gly Ile Lys Val Leu Lys Ile Glu Gly Phe 100 105 110

Glu Ala Asp Asp Ile Ile Ala Thr Leu Ser Lys Lys Phe Glu Ser Asp 115 120 125

Phe Glu Lys Val Asn Ile Ile Thr Gly Asp Lys Asp Leu Leu Gln Leu 130 140

Val Ser Asp Lys Val Phe Val Trp Arg Val Glu Arg Gly Ile Thr Asp 145 150 155 160

Leu Val Leu Tyr Asp Arg Asn Lys Val Ile Glu Lys Tyr Gly Ile Tyr
165 170 175

Pro Glu Gln Phe Lys Asp Tyr Leu Ser Leu Val Gly Asp Gln Ile Asp 180 185 190

Asn Ile Pro Gly Val Lys Gly Ile Gly Lys Lys Thr Ala Val Ser Leu 195 200 205

Leu Lys Lys Tyr Asn Ser Leu Glu Asn Val Leu Lys Asn Ile Asn Leu 210 215 220

Leu Thr Glu Lys Leu Arg Arg Leu Leu Glu Asp Ser Lys Glu Asp Leu 225 230 235 240

- Gln Lys Ser Ile Glu Leu Val Glu Leu Ile Tyr Asp Val Pro Met Asp 250 Val Glu Lys Asp Glu Ile Ile Tyr Arg Gly Tyr Asn Pro Asp Lys Leu Leu Lys Val Leu Lys Lys Tyr Glu Phe Ser Ser Ile Ile Lys Glu Leu 280 Asn Leu Gln Glu Lys Leu Glu Lys Glu Tyr Ile Leu Val Asp Asn Glu 295 Asp Lys Leu Lys Lys Leu Ala Glu Glu Ile Glu Lys Tyr Lys Thr Phe 310 Ser Ile Asp Thr Glu Thr Thr Ser Leu Asp Pro Phe Glu Ala Lys Leu 325 Val Gly Ile Ser Ile Ser Thr Met Glu Gly Lys Ala Tyr Tyr Ile Pro Val Ser His Phe Gly Ala Lys Asn Ile Ser Lys Ser Leu Ile Asp Lys Phe Leu Lys Gln Ile Leu Gln Glu Lys Asp Tyr Asn Ile Val Gly Gln Asn Leu Lys Phe Asp Tyr Glu Ile Phe Lys Ser Met Gly Phe Ser Pro Asn Val Pro His Phe Asp Thr Met Ile Ala Ala Tyr Leu Leu Asn Pro Asp Glu Lys Arg Phe Asn Leu Glu Glu Leu Ser Leu Lys Tyr Leu Gly 425 Tyr Lys Met Ile Ser Phe Asp Glu Leu Val Asn Glu Asn Val Pro Leu 435 Phe Gly Asn Asp Phe Ser Tyr Val Pro Leu Glu Arg Ala Val Glu Tyr
- Arg Lys Ile Tyr Glu Asn Glu Met Glu Lys Leu Phe Tyr Glu 1le Glu 485 490 495

Ser Cys Glu Asp Ala Asp Val Thr Tyr Arg Ile Phe Arg Lys Leu Gly

- Met Pro Leu Ile Asp Val Leu Ser Glu Met Glu Leu Asn Gly Val Tyr
  500 505 510
- Phe Asp Glu Glu Tyr Leu Lys Glu Leu Ser Lys Lys Tyr Gln Glu Lys 515 520 525
- Met Asp Gly Ile Lys Glu Lys Val Phe Glu Ile Ala Gly Glu Thr Phe 530 535 540

Asn 545	Leu	Asn	Ser	Ser	550	GIn	val	Ala	Tyr	555	Leu	Phe	Glu	Lys	Leu 560		
Asn	Ile	Ala	Pro	Tyr 565	Lys	Lys	Thr	Ala	Thr 570	Gly	Lys	Phe	Ser	Thr 575	Asn	٠	
Ala	Glu	Val	Leu 580	Glu	Glu	Leu	Ser	Lys 585	Glu	His	Glu	Ile	Ala 590	Lys	Leu		
Leu	Leu	Glu 595	Tyr	Arg	Lys	Tyr	Gln 600	Lys	Leu	Lys	Ser	Thr 605	Tyr	Ile	Asp		
Ser	Ile 610	Pro	Leu	Ser	Ile	Asn 615	Arg	Lys	Thr	Asn	Arg 620	Val	His	Thr	Thr		
Phe 625	His	Gln	Thr	Gly	Thr 630	Ser	Thr	Gly	Arg	Leu 635	Ser	Ser	Ser	Asn	Pro 640		
Asn	Leu	Gln	Asn	Leu 645	Pro	Thr	Arg	Ser	Glu 650	Glu	Gly	Lys	Glu	Ile 655	Arg		
Lys	Ala	Val	Arg 660	Pro	Gln	Arg	Gln	Asp 665	Trp	Trp	Ile	Leu	Gly 670	Ala	Asp		
Tyr	Ser	Gln 675	Ile	Glu	Leu	Arg	Val 680	Leu	Ala	His	Val	Ser 685	Lys	Asp	Glu		
Asn	Leu 690	Leu	Lys	Ala	Phe	Lys 695	Glu	Asp	Leu	Asp	Ile 700	His	Thr	lle	Thr		
Ala 705	Ala	Lys	Ile	Phe	Gly 710	Val	Ser	Glu	Met	Phe 715	Val	Ser	Glu	Gln	Met 720		
Arg	Arg	Val	Gly	Lys 725	Met	Val	Asn	Phe	Ala 730	Ile	Ile	Tyr	Gly	Val 735	Ser		
Pro	Tyr	Cly	Leu 740	Ser	Lys	Arg	Ile	Gly 745		Ser	Val	Ser	Glu 750	Thr	Lys		
Lys	Ile	Ile 755	Asp	Asn	Tyr	Phe	Arg 760	Tyr	Tyr	Lys	Gly	Val 765	Phe	Glu	Tyr		
Leu	Lys 770	Arg	Met	Lys	Asp	Glu 775	Ala	Arg	Lys	Lys	Gly 780	Tyr	Val	Thr	Thr		
Leu 785	Phe	Gly	Arg	Arg	Arg 790	Tyr	Ile	Pro	Gln	Leu 795	Arg	Ser	Lys	Asn	Gly 800		
Asn	Arg	Val	Gln	Glu 805	Gly	Glu	Arg	Ile	Ala 810		Asn	Thr	Pro	71e 815	Gln		
Gly	Thr	Ala	Ala 820	_	Ile	Ile	Lys	Ile 825		Met	Ile	Asn	11e 830		Asn		
Arg	Leu	Lys 835	-	Glu	Asn	Leu	Arg 840	Ser	Lys	Met	Ile	Leu 845	Gln	Val	His		

WO 92/06200 PCT/US91/07035

-157-

Asp Glu Leu Val Phe Glu Val Pro Asp Asn Glu Leu Glu Ile Val Lys 850 855 860

Asp Leu Val Arg Asp Glu Met Glu Asn Ala Val Lys Leu Asp Val Pro 865 870 880

Leu Lys Val Asp Val Tyr Tyr Gly Lys Glu Trp Glu 885 890

- (2) INFORMATION FOR SEQ ID NO:13:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 33 nucleotides
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA probe BW33
  - (iii) HYPOTHETICAL: NO
  - (iv) ANTI-SENSE: NO
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GATCGCTGCG CGTAACCACC ACACCCGCCG CGC

33

- (2) INFORMATION FOR SEQ ID NO:14:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 30 nucleotides
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA primer BW37
  - (iii) HYPOTHETICAL: NO
  - (iv) ANTI-SENSE: NO
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCGCTAGGGC GCTGGCAAGT GTAGCGGTCA

(2) INFO	RMATION FOR SEQ ID NO:15:
<b>(i)</b>	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 4 amino acids  (B) TYPE: amino acid  (D) TOPOLOGY: linear
(ii)	MOLECULE TYPE: peptide
(iii)	HYPOTHETICAL: YES
(iv)	ANTI-SENSE: NO
(ix)	FEATURE:  (A) NAME/KEY: Peptide  (B) LOCATION: 14  (D) OTHER INFORMATION: /label= Xaa  /note= "Xaa = Val or Thr"
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:15:
Ala 1	Xaa Tyr Gly
(2) INFO	RMATION FOR SEQ ID NO:16:
(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 5 amino acids  (B) TYPE: amino acid  (D) TOPOLOGY: linear
(ii)	MOLECULE TYPE: peptide
(iii)	HYPOTHETICAL: NO
(iv)	ANTI-SENSE: NO
(v)	FRAGMENT TYPE: internal
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:16:
His 1	Glu Ala Tyr Gly 5

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 5 amino acids

- (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

His Glu Ala Tyr Glu
1 5

- (2) INFORMATION FOR SEQ ID NO:18:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 4 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: peptide
  - (iii) HYPOTHETICAL: NO
  - (iv) ANTI-SENSE: NO
  - (v) FRAGMENT TYPE: internal
  - (ix) FEATURE:
    - (A) NAME/KEY: Peptide
    - (B) LOCATION: 1..4
    - (D) OTHER INFORMATION: /label= Xaa /note= "Xaa = Leu or Ile"
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Xaa Leu Glu Thr l

- (2) INFORMATION FOR SEQ ID NO:19:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 7 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear

```
(ii) MOLECULE TYPE: peptide
   (iii) HYPOTHETICAL: NO
   (iv) ANTI-SENSE: NO
    (v) FRAGMENT TYPE: internal
   (ix) FEATURE:
          (A) NAME/KEY: Peptide
          (B) LOCATION: 1..7
          (D) OTHER INFORMATION: /label= Xaa
                 /note= "Xaa = Leu or Ile"
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
    Xaa Leu Glu Thr Tyr Lys Ala
(2) INFORMATION FOR SEQ ID NO:20:
    (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 7 amino acids
          (B) TYPE: amino acid
         (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: peptide
   (iii) HYPOTHETICAL: NO
   (iv) ANTI-SENSE: NO
    (v) FRAGMENT TYPE: internal
   (ix) FEATURE:
         (A) NAME/KEY: Peptide
         (B) LOCATION: 1..7
         (D) OTHER INFORMATION: /label= Xaa1-4
                 /note= "Xaal = Ile or Leu or Ala; Xaa2-4, each =
                 any amino acid"
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
    Xaa Xaa Xaa Tyr Lys Ala
                     5
```

121	INFORMATION	FOR	SEO	TD	NO:21:
( Z )	TMLOMMATION	LOI	Y LQ	10	110.21.

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA primer MK61
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

## AGGACTACAA CTGCCACACA CC

22

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 38 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA primer RA01
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

## CGAGGCGCGC CAGCCCCAGG AGATCTACCA GCTCCTTG

38

## (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA primer DG29
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
AGCTTATGTC TCCAAAAGCT	20
	;
(2) INFORMATION FOR SEQ ID NO:24:	ź
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 16 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: DNA primer DG30	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
AGCTTTTGGA GACATA	16
(2) INFORMATION FOR SEQ ID NO:25:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 25 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: DNA primer PL10	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
GGCGTACCTT TGTCTCACGG GCAAC	25
(2) INFORMATION FOR SEQ ID NO:26:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 28 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	

		•	
	(ii)	MOLECULE TYPE: DNA primer FL63	
	(iii)	HYPOTHETICAL: NO	
ŧ	(iv)	ANTI-SENSE; NO	
·	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:26:	
	GATAAAGG	CA TGCTTCAGCT TGTGAACG	28
	(2) INFO	RMATION FOR SEQ ID NO:27:	
	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 27 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA primer FL69	
	(iii)	HYPOTHETICAL: NO	
	(iv)	ANTI-SENSE: NO	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:27:	
	TGTACTTC	TC TAGAAGCTGA ACAGCAG	27
	(2) INFO	RMATION FOR SEQ ID NO:28:	
	(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 36 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA primer FL64	
,	(iii)	HYPOTHETICAL: NO	
-	(iv)	ANTI-SENSE: NO	
a .	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:28:	

CTGAAGCATG TCTTTGTCAC CGGTTACTAT CAATAT

(iv) ANTI-SENSE: NO

(2) INFORMATION FOR SEQ ID NO:29:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 18 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	1
(ii) MOLECULE TYPE: DNA primer FL65	
(iii) HYPOTHETICAL: NO	•
(iv) ANTI-SENSE: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	
TAGTAACCGG TGACAAAG	18
(2) INFORMATION FOR SEQ ID NO:30:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 31 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA primer FL66	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	
· · · · · · · · · · · · · · · · · · ·	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:	21
CTATGCCATG GATAGATCGC TTTCTACTTC C	31
(2) INFORMATION FOR SEQ ID NO:31:	,
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 31 nucleotides</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: DNA primer FL67	
(iii) HYPOTHETICAL: NO	

(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:31:	
CAAGCCCA	TG GAAACTTACA AGGCTCAAAG A	31
(2) INFO	RMATION FOR SEQ ID NO:32:	
(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 49 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: DNA primer TZA292	
(iii)	HYPOTHETICAL: NO	
(iv)	ANTI-SENSE: NO	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:32:	
GTCGGCAT	AT GGCTCCTGCT CCTCTTGAGG AGGCCCCCTG GCCCCCGCC	49
(2) INFO	RMATION FOR SEQ ID NO:33:	
(i)	SEQUENCE CHARACTERISTICS:  (A) LENGTH: 37 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: DNA primer TZR01	
(iii)	HYPOTHETICAL: NO	
(iv)	ANTI-SENSE: NO	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:33:	
GACGCAGA'	TC TCAGCCCTTG GCGGAAAGCC AGTCCTC	37

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA primer TSA288	
(iii) HYPOTHETICAL: NO	
(iv) ANTI-SENSE: NO	3
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	ą
GTCGGCATAT GGCTCCTAAA GAAGCTGAGG AGGCCCCCTG GCCCCCGCC	49 .
(2) INFORMATION FOR SEQ ID NO:35:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 37 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA primer TSR01	
(iii) HYPOTHETICAL: NO	_
(iv) ANTI-SENSE: NO	•
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
GACGCAGATC TCAGGCCTTG GCGGAAAGCC AGTCCTC	37
(2) INFORMATION FOR SEQ ID NO:36:	
(i) SEQUENCE CHARACTERISTICS:  (A) LENGTH: 41 nucleotides  (B) TYPE: nucleic acid  (C) STRANDEDNESS: single  (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA primer DG122	
(iii) HYPOTHETICAL: NO	•
(iv) ANTI-SENSE: NO	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:	• ·
CCTCTAAACG GCAGATCTGA TATCAACCCT TGGCGGAAAG C	41

(2)	INFORMATION	FOR	SEQ	ID	NO:37:
-----	-------------	-----	-----	----	--------

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 48 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA primer TAFI285
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GTCGGCATAT GATTAAAGAA CTTAATTTAC AAGAAAAATT AGAAAAGG

48

- (2) INFORMATION FOR SEQ ID NO:38:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 46 nucleotides
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA primer TAFR01
  - (iii) HYPOTHETICAL: NO
  - (iv) ANTI-SENSE: NO
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCTTTACCCC AGGATCCTCA TTCCCACTCT TTTCCATAAT AAACAT

#### WHAT IS CLAIMED IS:

- A recombinant thermostable DNA polymerase enzyme which exhibits altered 5' to 3' exonuclease activity from that of its native DNA polymerase.
- 2. The recombinant thermostable DNA polymerase enzyme of claim 1 wherein a greater amount of 5' to 3' exonuclease activity is exhibited than that of the native DNA polymerase.
- 3. The recombinant thermostable DNA polymerase enzyme of claim 2 comprising the amino acid sequence A(X)YG wherein X is V or T (SEQ ID NO:15), and/or the amino acid sequence X<sub>A</sub>X<sub>3</sub>YKA wherein X<sub>A</sub> is I, L or A and X<sub>3</sub> is any sequence of three amino acids (SEQ ID NO:20).
- 4. The recombinant thermostable DNA polymerase enzyme of claim 1 wherein a lesser amount of 5' to 3' exonuclease activity is exhibited than that of the native DNA polymerase.
- 5. The recombinant thermostable DNA polymerase enzyme of claim 4 which in its native form comprises the amino acid sequence A(X)YG wherein X is V or T (SEQ ID NO:15), said amino acid sequence being mutated or deleted in said recombinant enzyme.
- 30 6. The recombinant thermostable DNA polymerase enzyme of claim 5 wherein G of SEQ ID NO:15 is mutated.
- 7. The recombinant thermostable DNA polymerase enzyme of claim 6 wherein G of SEQ ID NO:15 is mutated to A.

WO 92/06200 PCT/US91/07035

-169-

8. The recombinant thermostable DNA polymerase enzyme of claim 4 which in its native form comprises the amino acid sequence HEAYG (SEQ ID NO:16), said amino acid sequence being mutated or deleted in said recombinant enzyme.

9. The recombinant thermostable DNA polymerase enzyme of claim 4 which in its native form comprises the amino acid sequence HEAYE (SEQ ID NO:17), said amino acid sequence being mutated or deleted in said recombinant enzyme.

10. The recombinant thermostable DNA polymerase enzyme of claim 4 which in its native form comprises the amino acid sequence XLET wherein X is L or I (SEQ ID NO:18), said amino acid sequence being mutated or deleted in said recombinant enzyme.

11. The recombinant thermostable DNA polymerase enzyme
20 of claim 4 selected from the group consisting of
mutant forms of <u>Thermus</u> species sps17, <u>Thermus</u>
species Z05, <u>Thermus</u> <u>aquaticus</u>, <u>Thermus</u>
thermophilus, <u>Thermosipho</u> <u>africanus</u> and <u>Thermotoga</u>
maritima.

25

5

10

12. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="mailto:thermus">Thermus</a> aguaticus comprising amino acids 77-832 of SEO ID NO:2.

30

13. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/jheps.comprising-amino-acids-47-832">https://doi.org/10.1001/jheps.comprising-amino-acids-47-832</a> of SEQ ID NO:2.

ġ

3

\*

5

10

15

17. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1016/jhear.2015/national.org/">Thermotoga maritima comprising amino acids 38-893 of SEQ ID NO:4.</a>

20

25

19. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermotoga maritima</u> comprising amino acids 74-893 of SEQ ID NO:4.

30

20. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/nc.2007/nc

PCT/US91/07035

-171-

5

ı

WO 92/06200

22. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/jhep-10.

10

23. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/jheps://do

15

24. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://docs.org/thermus">Thermus</a> species sps17 comprising amino acids 152-830 of SEQ ID NO:6.

20

25. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://docs.org/thermus">Thermus</a> species sps17 comprising amino acids 200-830 of SEQ ID NO:6.

25

26. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="mailto:Thermus">Thermus</a> species sps17 comprising amino acids 288-830 of SEO ID NO:6.

30

27. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/nc

1

)

28. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="Thermus">Thermus</a> species Z05 comprising amino acids 78-834 of SEQ ID NO:8.

5

29. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/nc.2015">Thermus</a> species Z05 comprising amino acids 156-834 of SEQ ID No:8.

10

30. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/nc.2015">Thermus</a> species Z05 comprising amino acids 204-834 of SEQ ID No:8.

15

31. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/jhep-10.

20

32. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10">Thermus thermophilus</a> comprising amino acids 47-834 of SEQ ID NO:10.

25

33. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10.10">Thermus thermophilus comprising amino acids 78-834</a> of SEQ ID NO:10.

30

34. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10/10.10">Thermus thermophilus comprising amino acids 156-834</a> of SEQ ID NO:10.

WO 92/06200 PCT/US91/07035

-173-

35. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10/10.10">Thermus thermophilus comprising amino acids 204-834</a> of SEQ ID NO:10.

5

ι

36. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10/10.10">Thermus thermophilus</a> comprising amino acids 292-834 of SEO ID NO:10.

10

15

38. The recombinant thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1016/jhep-10.1016/">Thermosipho africanus comprising amino acids 94-892 of SEQ ID NO:12.</a>

20

25

30

ı

,

)

3

- 42. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1011/jhap.no.1011">Thermus aquaticus</a>, said DNA sequence comprising nucleotides 229-2499 of SEQ ID NO:1.
- 43. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus aquaticus</u>, said DNA sequence comprising nucleotides 139-2499 of SEQ ID NO:1.
- 44. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus aquaticus</u>, said DNA sequence comprising nucleotides 463-2499 of SEQ ID NO:1.
- 45. A DNA sequence which encodes a thermostable DNA
  20 polymerase enzyme of claim 11 wherein said enzyme
  is a mutant form of <u>Thermus aquaticus</u>, said DNA
  sequence comprising nucleotides 607-2499 of SEQ ID
  NO:1.
- 25 46. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus aquaticus</u>, said DNA sequence comprising nucleotides 868-2499 of SEQ ID NO:1.

30

35

- 48. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1016/journal.com/">Thermotoga maritima</a>, said DNA sequence comprising nucleotides 61-2682 of SEQ ID NO:3.
- 49. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10/10.10-10.
- 50. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/jhaps.com/rising-nucleotides-418-2682">https://doi.org/10.1001/jhaps.com/rising-nucleotides-418-2682</a> of SEQ ID NO:3.
- 51. A DNA sequence which encodes a thermostable DNA
  20 polymerase enzyme of claim 11 wherein said enzyme
  is a mutant form of Thermotoga maritima, said DNA
  sequence comprising nucleotides 850-2682 of SEQ ID
  NO:3.
- 25 52. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species sps17, said DNA sequence comprising nucleotides 130-2493 of SEQ ID NO:5.

53. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species sps17, said DNA sequence comprising nucleotides 220-2493 of SEQ ID

35 NO:5.

30

ı

١.

5

J

1

)

٠

- 54. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species sps17, said DNA sequence comprising nucleotides 454-2493 of SEQ ID NO:5.
- 55. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species sps17, said DNA sequence comprising nucleotides 598-2493 of SEQ ID NO:5.
- 56. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species sps17, said DNA sequence comprising nucleotides 862-2493 of SEQ ID NO:5.
- 57. A DNA sequence which encodes a thermostable DNA
  20 polymerase enzyme of claim 11 wherein said enzyme
  is a mutant form of Thermus species Z05, said DNA
  sequence comprising nucleotides 139-2505 of SEQ ID
  NO:7.
- 25 58. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species Z05, said DNA sequence comprising nucleotides 232-2505 of SEQ ID NO:7.

30

35

59. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species 205, said DNA sequence comprising nucleotides 476-2505 of SEQ ID NO:7.

10

- 60. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.1001/jheps.com/">Thermus</a> species 205, said DNA sequence comprising nucleotides 610-2505 of SEQ ID NO:7.
- 61. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus</u> species Z05, said DNA sequence comprising nucleotides 874-2505 of SEQ ID NO:7.
- 62. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus thermophilus</u>, said DNA sequence comprising nucleotides 139-2505 of SEQ ID NO:9.
- 63. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10">Thermus thermophilus</a>, said DNA sequence comprising nucleotides 232-2505 of SEQ ID NO:9.
- 25 64. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus thermophilus</u>, said DNA sequence comprising nucleotides 466-2505 of SEQ ID NO:9.

30

35

65. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus thermophilus</u>, said DNA sequence comprising nucleotides 610-2505 of SEQ ID NO:9.

10

ĵ

j

3

¥

- 66. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermus thermophilus</u>, said DNA sequence comprising nucleotides 874-2505 of SEQ ID NO:9.
- 67. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="Thermosipho africanus">Thermosipho africanus</a>, said DNA sequence comprising nucleotides 112-2679 of SEQ ID NO:11.
- 69. A DNA sequence which encodes a thermostable DNA
  20 polymerase enzyme of claim 11 wherein said enzyme
  is a mutant form of Thermosipho africanus, said DNA
  sequence comprising nucleotides 418-2679 of SEQ ID
  NO:11.
- 25 70. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <a href="https://doi.org/10.10/10.10-2679">Thermosipho africanus</a>, said DNA sequence comprising nucleotides 610-2679 of SEQ ID NO:11.

30

35

71. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 11 wherein said enzyme is a mutant form of <u>Thermosipho africanus</u>, said DNA sequence comprising nucleotides 853-2679 of SEQ ID NO:11.

- 72. A DNA sequence which encodes a thermostable DNA polymerase enzyme of claim 3.
- 73. A DNA sequence which encodes a thermostable DNA polymerase enzyme of any of claim 5 through 10.
  - 74. A recombinant DNA vector comprising the DNA sequence of any of claims 42 through 73.
- 10 75. A recombinant host cell transformed with the vector of claim 74.

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 91/07035

		CT MATTER (if several classification		
According : Int.Cl		Classification (JPC) or to both Nations C 12 N 15/54 C		/21
II. FIELDS	SEARCHED			<del> </del>
		Minimum Doc	umentation Searched <sup>7</sup>	
Classificati	on System		Classification Symbols	
Int.Cl	.5	C 12 N		
		Documentation Searched of to the Extent that such Docume	her than Minimum Documentation hts are Included in the Fields Searched <sup>8</sup>	
		:		
<u> </u>		D TO BE RELEVANT <sup>9</sup>		Relevant to Claim No.13
Category °	Citation of Do	cument, 11 with indication, where appro	opriate, of the relevant passages 14	Relevant to Claim No.13
Р,Х	(Bethe "Chara activi	SEB Journal, volume 5 sda, MD, US) R.D. Abr cterization of the 5' ty of thermus aquatic 437, abstract no. 386	amson et al.: -3'exonuclease us DNA polymerase",	1,4
P,Y				11-16
P,X	1991, 5 7-25; j	109944 (CETUS CORP.) see page 13, lines 28 page 14, line 34 - pa (cited in the applica	-34; page 14, lines ge 16, line 15: claims	1,4
"A" docucons "E" earlii filin "L" docucons "O" docuconte "P" docu	idered to be of particular document but publis godate ment which may throw his cited to establish toon or other special realment referring to an or means ment published prior to than the priority date ICATION	eral state of the art which is not lar relevance shed on or after the international doubts on priority claim(s) or the publication date of another uson (as specified) ral disclosure, use, exhibition or the international filing date but claimed	To later document published after the internal or priority date and not in conflict with the cited to understand the principle or theory invention  "X" document of particular relevance; the claim cannot be considered novel or cannot be considered novel or cannot be considered novel or cannot be considered to involve an inventive step  "Y" document of particular relevance; the claim cannot be considered to involve an inventive document is combined with one or more of ments, such combination being obvious to in the art.  "A" document member of the same patent fame.  Date of Mailing of this International Sear.	the application but y underlying the winderlying the med invention med invention the the there were the there such docu- a person skilled will the the the the the the the the there were the the there were the there were the there were the there were the the there were the there were the there were the there were the the there were the there were the there were the there were the the there were the there were the there were the there were the the there were the there were the there were the there were the the there were the there were the there were the there were the the there were the there were the there were the there were the the there were the there were the there were the there were the the there were the there were the the there were the t
International	Searching Authority		Signature of Authorized Officer	
	EUROPEA	N PATENT OFFICE	Mme N. KUIPER	wife

III. DOCUMEN		/US 91/07035
	ITS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)	
Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Ρ,Υ	· · · · · · · · · · · · · · · · · · ·	1-11,17 -21,32- 36,42- 51,62- 66,72- 75
γ,Υ	WO,A,9109950 (CETUS CORP.) 11 July 1991, see claims 3,4 (cited in the application)	1-7,9, 11,32- 36,62- 66,72- 75
, X , Y	WO,A,9102090 (PROMEGA CORP.) 21 February 1991, see claim 1; page 6, line 34 - page 7, line 7	1,4,5, 11,14- 16,73 6-10,12 ,13,17- 72,74,
Y	The Journal of Biological Chemistry, volume 264, no. 11, 15 April 1989, Am. Soc. for Biochemist. and Molecular Biology, Inc. (US) F.C: Lawyer et al.: "Isolation, characterization, and expression in Eschericia coli of the DNA polymerase gene from thermus aquaticus", pages 6427-6437, see figure 2; page 6432, lines 25-30 (cited in the application)	1-8,11- 16,42- 46,72 - 75
Y	Cell, volume 59, no. 1, 6 October 1989, Cell Press (MA, US) A. Bernad et al.: "A conserved 3'-5' exonuclease active site in prokarytic and eukarytic DNA polymerases", pages 219-228, see page 224, lines 6-10, line 21 - page 225, line 14; figure 5	1,4-75
Y	Proc. Natl. Acad. Sci., volume 86, no. 12, June 1989, Biochemistry (US) M.C. Leavitt et al.: "T5 DNA polymerase: structural-functonal relationships to other DNA polymerases", pages 4465-4469, see figure 4; page 4468, lines 13-35	1,4-75

Form PCT/ISA/210 (cotrs sheet) (January 1985)

International Application No

Page 3 PCT/US 91/07035

III DOCUMEN	International Application No PCT	<u>/US 91/07035</u>
	(CONTINUED FROM THE SECOND SHEET)	
Category °	Citation of Document, with Indication, where appropriate, of the relevant passages	Relevant to Claim No
A	Chemical Abstracts, volume 93, no. 5, 4 August 1980 (Columbus, Ohio, US) A.S. Kaledin et al.: "Isolation and properties of DNA polymerase from extremal thermophylic bacteria Thermus aquaticus YT-1", see page 377, abstract 40169p, & Biokhimiya (Moscow) 1980, 45(4), 644-51	1,4
A .	Chemical Abstracts, volume 85, no. 21, 22 November 1976, (Columbus, Ohio, US) A. Chien et al.: "Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus", see page 180, abstract 155559t, & J. Bacteriol. 1976, 127(3), 550-7	1,4
	·	

Form PCT/ISA/210 (extra sheet) (James 1985)

# ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

US 9107035

SA 52103

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 13/12/91

The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)  WO-A- 9109950		Publication date
WO-A- 9109944	11-07-91			11-07-91
₩0-A- 9109950	11-07-91	WO-A-	9109944	11-07-91
WO-A- 9102090	21-02-91	AU-A-	6341190	11-03-91
			_	
	•			
				·.
•				
			,	

PORM P0479

E For more details about this annex : see Official Journal of the European Patent Office, No. 12/82